RC22481 (W0206-033) June 10, 2002
Computer Science

IBM Research Report

Biometrics 101

Rudolf M. Bolle, Jonathan Connell, Sharathchandra Pankanti,
Nalini K. Ratha, Andrew W. Senior
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

Research Division
Almaden - Austin - Beijing - Delhi - Haifa - India - T. J. Watson - Tokyo - Zurich

Biometrics 101

Version 6.05

R.M. Bolle, J.H. Connell, S. Pankanti, N.K. Ratha and A.W. Senior
IBM Thomas J. Watson Research Center

Abstract

There is much interest in the use of biometrics for *verification*, *identification*, and "*screening*" applications, collectively called *biometric authentication*. This interest has been heightened because of the threat of terrorism. Biometric authentication systems offer advantages over systems based on knowledge or possession such as unsupervised (legacy) authentication systems based on password/PIN and supervised (legacy) authentication systems based on driver's licences and passports. The most important advantage is increased security: when a person is authenticated based on a biometric, the probability that this person is the originally enrolled person can be statistically estimated or computed in some other way. When a person is authenticated based on a password or even based on human observation, no such probabilities can be determined. Of course, the mere capability to compute this probability is not sufficient, what is needed is that the probability of correct authentication is high and the error probabilities are low. Achieving this probabilistic linking by introducing biometrics in authentication systems brings along many design choices and may introduce additional security loopholes. This document studies the many aspects of biometric applications that are an issue even before a particular biometrics has been selected; it further studies many issues that are associated with the currently popular biometric identifiers, namely, finger, face, voice, iris, hand (geometry) and signature.

Contents

List of Figures

List of Tables

Biometrics 101

Preface

Biometrics is a fascinating area that draws from many disciplines, but pattern recognition is the main discipline that concern itself with biometrics. Biometrics is in the process of forming a community in itself. Other disciplines that should be more involved in biometrics are computer security and user interfaces, or more generally the area of human interaction technologies.

Various areas from pattern recognition, image processing, computer vision, signal processing (speech), document analysis are concerned with developing matching technologies for the various biometrics. Because this is an area of pattern recognition, it brings along by necessity the ideas of probability theory and the *a priori* fact that a person or identity can only be authenticated subject to the probability of making mistakes: falsely authenticating (accepting) a possible intruder, or falsely rejecting a genuine user. Because of this, the computer security community may not feel that biometrics is in their charter. We hope that a consequence of this document is that the security and biometrics areas open more of a dialogue. It only appears to be a matter of time till authentication protocols as a security service, as defined by information security analysts, will include biometrics.

Biometrics is an area of many facets, many of which we have attempted to review, or at least to explain in this document. Our intent was to allow the question: "Which biometric is best?" to be answered in a better-informed, detailed way using this document. While the answer may be only tentative, and surely not definitive, at least the question can be phrased a little better by asking: "Which biometric is best, given that the requirements for the application are specified as follows ..." Whether we have succeeded in this goal, we will of course not know without feedback from the readers. Therefore, the authors would like to invite you to contact them about any remaining unanswered concerns.

We could have continued writing about biometrics for quite a while longer however we somewhat arbitrarily decided to finish by June 5, 2002. This was the date of the *IBM Biometric Council* meeting at the IBM Thomas J. Watson Research Center. Therefore, the version number of this document is simply **Version 6.05**. We cannot guarantee that every piece of information contained in this document is 100% accurate; we have undertaken every effort to be as precise as possible.

The authors would like to acknowledge the support of IBM Global Services and IBM Research; we hope that it will help IGS and their customers in making well-informed decisions about biometric solutions. We also acknowledge John McKeon, co-chair of the Biometrics Council, for his valuable feedback on this document.

Hawthorne, NY
June 5, 2002

Biometrics 101

R.M. Bolle, J.H. Connell, S. Pankanti, N.K. Ratha and A.W. Senior
IBM Thomas J. Watson Research Center

1 Introduction

A biometric authentication system has to satisfy requirements that are often contradicting, most notably such a system has to guarantee safety without compromising too much on convenience and has to achieve this cost effectively. The desired recognition accuracy of a system is typically specified in terms of *quantitative* parameters such as error rates, which of course all need to be as low as possible. There are more *qualitative* properties like levels of security, convenience and privacy that also need to be satisfied in addition.

A problem is that the literature on biometrics is confusing and obscure when it comes to error numbers of the various biometrics. Only sparse and incomplete information and often contradictory information is in general available. Designing a biometric system hence brings along many questions: Can the quoted error numbers be trusted? How do we even measure the recognition accuracy of a biometric authentication system once it is installed? What exactly is security when it comes to trusting people instead of information? Do we create new security holes because of the use of biometrics? The list of items and questions goes on and on.

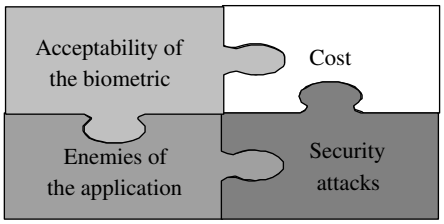


Figure 1: Designing a biometric authentication system involves many pieces of a puzzle.

The many aspects of biometric authentication form a difficult puzzle (Figure 1). This document is intended to introduce and explain many of the issues and to answer many of the questions.

1.1 Biometrics

In the modern networked society, there is an ever-growing need to determine or verify the identity of a person. Where authorization is necessary for any action, be it picking up a child from daycare or boarding an aircraft, authorization is almost always vested in a single individual or in a group of individuals. Identity verification becomes a challenging task when it has to be automated with high accuracy and hence with reliable non-reputability.

There are a number of methods adopted by society to verify identity of (authenticate) a person. These *traditional* methods of identity verification can be grouped into three classes [106]:

1. *Possession:* These are physical possessions such as keys, passports, and smart cards; in this document we *explicitly* include logical possessions like user IDs, account numbers.

2. *Knowledge:* Pieces of information (passwords) that are supposed to be kept secret and only are known to the rightful person. Other things that might not be *that secret*, such as *mother's maiden name*, are also considered knowledge.

3. *Biometrics:* Physiological and behavioral appearances or characteristics of individuals that distinguish one person from the next. These are characteristics of the human body and human actions that differentiate people from each other.

Often, the above three identification methods are used in combination, especially when it comes to automated authentication. A password plus a user ID is a knowledge method of identification; an ATM card is a possession that requires knowledge to carry out a transaction; a passport is a possession that requires (still human) biometric verification.

Biometrics is the science of identifying or verifying the identity of a person based on physiological or behavioral characteristics. Physiological characteristics include fingerprints and facial image; behavioral characteristics include signature and voice, see Miller [106]. Physiological biometrics, like fingerprints, are physical characteristics that are not influenced by emotional state. Behavioral biometrics, like signature, on the other hand, are believed to depend more on the state of mind of the subject and are learned or acquired over time.

Physiological	**Behavioral**
Face	Signature
Finger	Voice
Hand geometry	
Iris	

Table 1: The six biometrics handles in this document.

Another way of distinguishing between physiological and behavioral biometrics is that physiological biometrics are rich enough that a snapshot suffices. Behavioral biometrics are weaker and need multiple samples over time to build up uniqueness. Some biometric identifiers, the ones handled more in depth in this document, are given in Table 1.

The issue whether a biometric property is physiological or behavioral is not really that important, this is simply a distinction proposed by Miller [106]. What is important are the necessary attributes of a biometric described by Clarke [31] that characterize a biometric. These include:

1. *Universality,* which means that every person should have the characteristic;

2. *Uniqueness,* which means that no two persons should be the same in terms of the characteristic;

3. *Permanence,* which means that the characteristic should be invariant over time,

4. *Collectability,* which means that the characteristic can be measured with some sensing device.

It is the combination of all these attributes that determine the effectiveness of a biometric and therefore the effectiveness of biometric authentication using a particular biometric in a particular application [70]. There

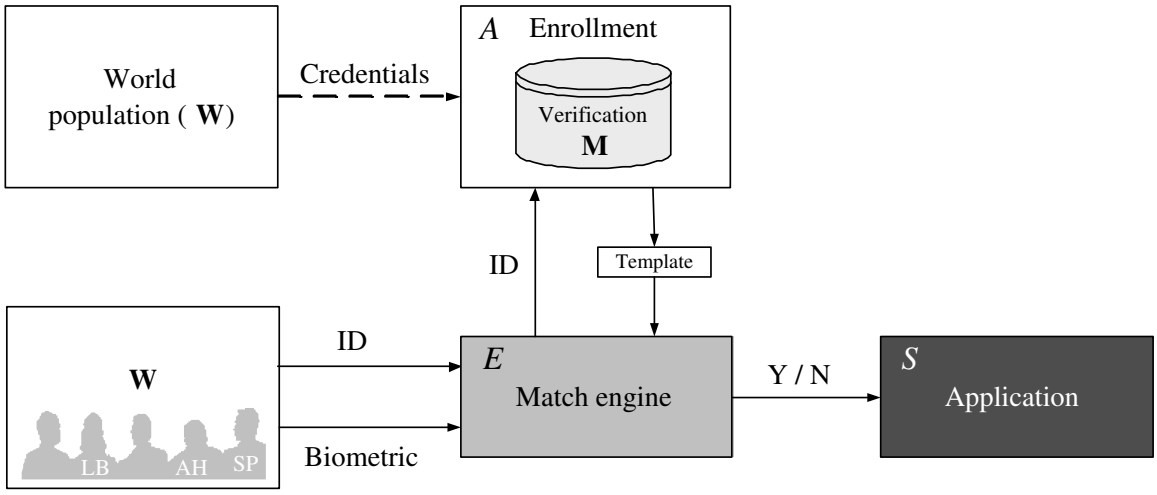


Figure 2: A biometric verification system, enrolling with a biometric and ID.

is really no biometric that satisfies these four properties to a satisfactory level simultaneously, especially if *acceptability* (acceptance of the use of a biometric by the user group and the public in general) is taken into account. This means that *any* biometric authentication solution is the result of many compromises related to many issues. Our goal is to view biometric authentication in light of these many issues.

1.2 Biometric authentication

Biometric authentication is about better securing applications or facilities (e.g., airports) against attacks at a certain cost. Biometrics is a fledgling technology area and there is a scattering of more or less successful trials around with various levels of security ("safety") of the application. Biometrics is also surrounded by much *hype* that we will attempt to untangle in this document.

Let us first introduce biometric authentication systems. Figure 2 shows the architecture of a system to implement *biometric verification,* that is, $1:1$ biometric matching: The top two boxes form the enrollment process A; the bottom three boxes perform the authentication process. There exists the (world) population $\mathbf{W}$ of human beings from which a portion is enrolled in a *verification database* $\mathbf{M}$ using enrollment process A (Figure 2). During enrollment a machine representation (*template*) of a person's biometric sample is computed; other credentials are also collected.

For authentication, the required credentials of an enrolled subject are presented, that is, an ID & input biometric, and a machine representation of the input biometric is computed. The *biometric match engine* E compares the stored template in $\mathbf{M}$ associated with the ID with the input biometric sample. The *degree of match* determines if the subject is eligible to access the application S (see Figure 2). Here the application could be, e.g., a travel application or a financial application. The application determines assets to be protected, *friends,* and *foes* in world population $\mathbf{W}$. In sum, a biometric verification system for application S at least consists of an *enrollment* system A, a verification database $\mathbf{M}$, and a matching engine E.

Biometric verification only requires $1:1$ matching. Another biometric function called *positive identification* refers to determining that a given individual is in some existing database $\mathbf{M}$. *Negative identification* is determining that a given individual is *not* in some "most wanted" database $\mathbf{N}$. Positive and negative identification amount to a similar search problem of database $\mathbf{M}$ and $\mathbf{N}$ of biometric templates, respectively. We

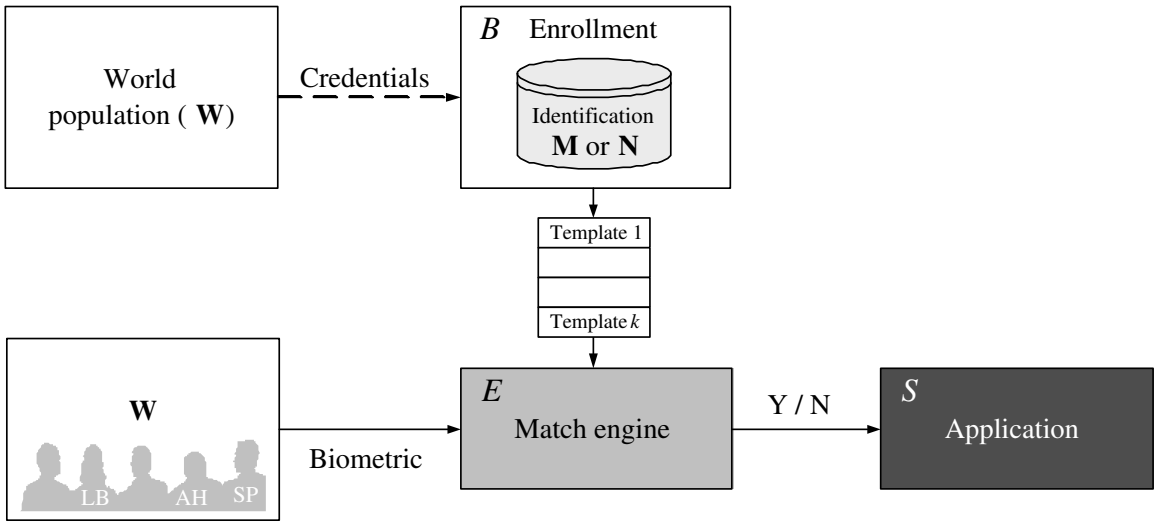


Figure 3: A biometric screening system, "enrolling" *only* with a biometric.

call negative identification "*biometric screening*" in this document.

Figure 3 shows a system to implement biometric identification. If there are n entities in the biometric database, this identification problem amounts to $1 : n$ biometric matching (or *searching*). Again, the top two boxes form the enrollment process B, with two possibilities:

1. There is the (world) population $\mathbf{W}$ from which a portion is enrolled in an *identification database* $\mathbf{M}$ using a registration (enrollment) process B based on a subject's credentials, where database $\mathbf{M}$ contains m entities. The enrollment process could be based on the same credentials as the verification system of Figure 2. Hence in that case we have $B = A$.

2. Again, a portion of the world population $\mathbf{W}$ is enrolled in a *screening database* $\mathbf{N}$, containing n entities. However, in this case a subject's credentials are negative aspects of the person, such as arrests.

The bottom three boxes of Figure 3 perform the (positive) identification or the screening, using databases $\mathbf{M}$ and $\mathbf{N}$, respectively.

Let us concentrate on screening. Pure biometric screening proceeds with *only* a biometric sample as input to the match engine E. Biometric screening then amounts to matching the input biometric to each of the n stored templates. In general, a list of templates, (Template$_1$, ..., Template$_k$), is returned. The list corresponds to a set of individuals that are biometrically similar to the input biometric. A subject is accepted to the application S if her/his biometric is *not* on the list (Figure 3); if the biometric is on the list, presumably other measures will be taken.

It is seen that the enrollment processes A and B are an integral and important part of a biometric authentication system, where the enrollment module is a separate process that needs to be secured too. A biometric verification system and screening system have different enrollment processes:

- Enrollment A for verification: The purpose of enrollment for verification is to construct a verification database $\mathbf{M}$ of m eligible members. It has to be somehow determined what makes a subject eligible for the application.

Biometrics samples and other credentials are stored in **M**.

- Enrollment *B* for screening: Collection of a screening database **N** of n ineligible members from a population **W**. Here it has to be somehow determined for what reasons a subject would be ineligible. (The rules *themselves* that are used to exclude individuals from an application is of course not a topic of this document.)

  Biometrics samples and other credentials are stored in **N**.

These selection processes can be based on information about the population **W** in the form of "ground truth", i.e., data documented by birth certificates, passports, etc. in legacy databases and criminal data documented in government databases. The processes A and B implement *enrollment policies*, defined in this document, as a set of rules.

Upon a request for authentication, determining whether a subject is eligible or not is accomplished through an *authentication protocol* as introduced in this document. These protocols involve credentials in the form of possessions, knowledge, and biometrics. An authentication protocol can demand verification based on (ID, biometric); an authentication protocol can require verification based on $(ID, biometric)$ while it also requires screening based on just a biometric sample $(biometric)$.

But before proceeding in more depth, we introduce the various aspects of biometrics a little further in Section 2. The reader may either continue reading to get a brief overview of biometric issues, or may skip to Section 3. In Section 2, we describe the organization of this document and in addition we introduce much of the additional terminology (beyond what is introduced in this section). Moreover, what we try to do is to give a summary of all the aspects of an end-to-end biometric authentication system.

2 Organization of this document & terminology

This document, and this section, is roughly divided up into three parts.

Part I *System basics* (Sections 2-4): This part introduces the authentication protocols for verification, identification, and screening; it shows how these protocols can be augmented with biometrics (see Section 3).

This part further introduces the most common biometrics (finger, face, voice, iris, hand, signature) in Section 4.

Part II *Performance and selection* (Sections 5-10): This part of the document describes issues that *need* to be understood in order to understand biometrics and biometric applications.

Part II defines the accuracy of biometric systems (Section 5) and other more qualitative properties of these systems (Section 6). Realistic error rates and "intrinsic error rates" are discussed in Sections 7 and 8. Application properties are discussed in Section 9, properties of the biometrics are discussed in Section 10.

Part III *Meta-issues and advanced topics* (Sections 11-15): These issues are *testing and training* (Section 11) and *enrollment* (Section 12), fundamental and very important issues in biometric authentication. Further biometric *points of attacks* are discussed (Section 13). Section 14 discusses the various types of integration that take place in biometric authentication installations.

Finally we discuss what is going on on the area of APIs, standards, and databases in Section 15. This information is of course quite time sensitive; much of the latest information can be found on the Biometrics Consortium Web site [9].

2.1 Part I: System basics

By *secure authentication* we mean: Checking a subject's *credentials* and *authorizing* the person to use services/ privileges or to access premises. This means somehow checking the credentials of a subject and making a *'YES/NO'* decision. Section 3 discusses what it really means to be authenticated, introducing and adopting some terminology from the field of computer security [118, 152].

Biometrics is just one of the three authentication *methods* ("credentials") that have evolved over the centuries [106]: *possession* (key, smart card), *knowledge* (password) and *appearance*. The latter method spans a range of personal traits, from face appearance to gait to voice. During authentication, a biometric measurement is *acquired* and matched against a stored *machine representation* of the *biometric identifier*. A *degree of match* (match probability) is computed while possession and password are checked by straight comparison.

Possessions include things like the ownership of accounts, i.e., often identifiers in the form of numbers; knowledge can include more or less personal and private information such as "*mother's maiden name*" and "*favorite color*."

Checking and matching of combinations of one or more authentication methods define different *authentication protocols*, see Section 3. An authentication protocol includes a set of credentials and in the case of biometric authentication this set includes particular biometrics, such as face and voice. An authentication protocol further prescribes how the different credentials are to be matched.

Before a subject can be authenticated, the subjects has the be *enrolled.* Even though this is an important aspect of biometric authentication, this topic is deferred to Part III.

Biometric authentication protocols

Biometric identifiers can be included in a secure authentication protocol (and hence have to be included in the enrollment policy, see Section 12).

Notably now, as already mentioned in Section 1, there are two diametrically opposite biometric authentication protocols for securing an application (see Figure 4).

(Positive) Identification: Perhaps the most elusive possibility of biometric technology is automated person identification. Here m identities (subjects) are enrolled in an application database **M**. The authentication protocol demands *only* the presentation of the biometric identifier to the authentication system and a subject is identified or not. This involves $1 : m$ matching (searching).

Negative identification, or screening, is the same problem as identification. With n identities (bad guys, crooks, terrorists) enrolled in a database **N** this amounts to $1 : n$ matching.

Another major use of negative identification is to prevent "double dipping" for social benefits, and to aid the unravelling of aliases in criminal cases.

Verification: The authentication protocol includes the presentation of possessions and/or knowledge besides just the presentation of a biometric identifier. These additional credentials uniquely define the enrolled identity in **M** and hence the associated biometric machine representation (template).

Verification is most often used for building access, and access to accounts, such as computer logon. It is also called $1 : 1$ matching.

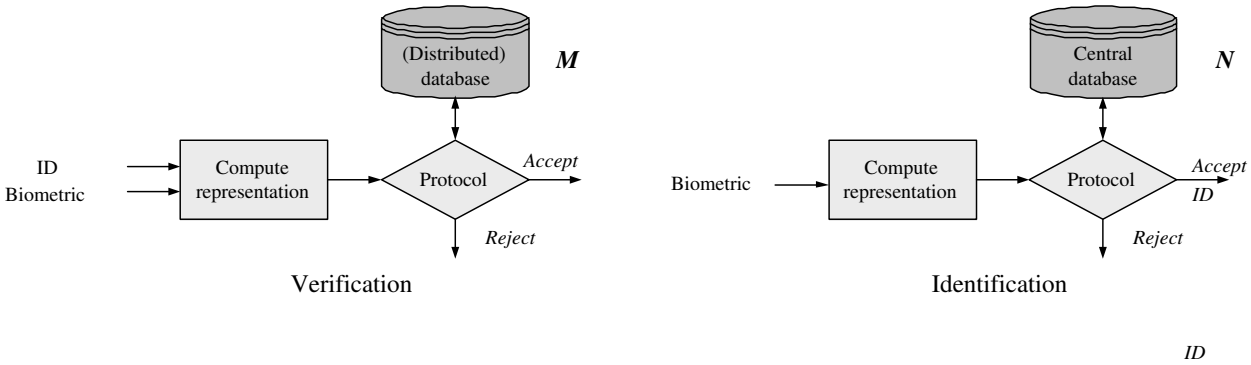


Figure 4: For verification, the identity of the subject is known, for identification, the identity of the subject is determined by matching multiple biometric representations.

Figure 4 shows the difference between verification and screening. Because the subject lays claim on an identity in the form of some account number, or in some other way, only $1 : 1$ matching is needed for verification; for identification, the identity of the subject has to be established through $1 : n$ biometric matching, with n the number of subjects in **N**.

Screening: In this document, with screening we mean the matching of multiple credentials (*tokens*) associated with a subject, among which are biometric identifiers, with information and biometric representations as stored in possibly large, often government databases (FBI, CIA, INS). This is sometimes called *negative identification* [168], i.e., assuring that a subject is *not* registered in a database.

This definition of screening is broader than the one in Section 1, where pure biometric screening is discussed.

Typically screening is done against databases that are collected using involuntary enrollment. Screening is just another form of authentication in that the protocol calls for checking many tokens, many of these of may be parametric and numeric such as names, passport numbers, etc. (see Figure 5).

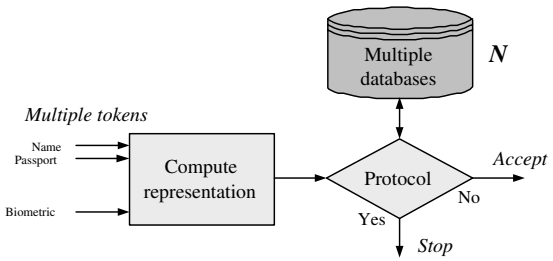


Figure 5: Screening is done by matching many pieces of information.

Continuity of identity: This is the problem of authenticating a subject and then, in a continuous mode by maximizing the confidence in authentication, ensuring authenticity of the subject over time given additional sensory data. For instance, is the person checked in at the gate the same one that boarded the airplane? This is a relatively unexplored research area but has its application in surveillance; issues around continuity of identity are discussed in Section 3.9.

Commonly used biometrics

Section 4 introduces and discusses the most commonly used biometrics in today's automated authentication systems. These biometric identifiers are: *fingerprint*; *face*; *voice*; *iris*; *signature* and *hand geometry* (Table 2).

Physiological	**Behavioral**
Face	Gait
Finger	Key strokes
Hand geometry	*Signature*
Iris	*Voice*
Retina	...
...	

Table 2: There are many human biometric characteristics.

The biometric iris is often confused with the biometric retina and sometimes called "eye biometric." In reality, iris and retina require quite different biometric sample acquisition. For obtaining a retina image, the pupil has to be diluted; for iris, images "reasonable sensing" environments suffice, as claimed by manufacturers. We describe these biometric identifiers in the framework of a pattern recognition system that detect a match, from input sensor to matching module that produces the *'Accept/Reject'* answer.

Table 2 lists two behavioral biometrics, *gait* and *keystrokes* that are in an early research stage. The *retina,* a physiological biometric, is only used in high security applications. These more esoteric biometric identifiers are further described in [70].

2.2 Part II: Performance and selection

A number of parameters (including *error rates*) need to be defined. These parameters could be associated with a particular biometric or could be associated with a particular biometric installation.

Part II first defines these parameters (Sections 5 and 6) and then we will give current results of error estimates for the six biometrics in Section 7. Section 8 deals with the so-called *intrinsic error* rates of biometrics. Further we discuss applications in Section 9 and give guidelines on how to select a biometric in Section 10.

Parameters and properties

There are a number of aspects of biometric authentication that have to be understood irrespective of the particular biometric, application and protocol. These are parameters and properties of the authentication system in Figure 3 that performs verification and/or screening (in general, we call this authentication):

We define concepts like False Accept *FA*rate and other parameters such as

Quantitative parameters: Quantitative, measurable parameters of a biometric application can be defined to express the "*accuracy*" of a biometric authentication application.

The definition of a *biometric match* can be statistically defined, therefore, there are precise definitions for parameters such as the probability of accepting intruders, *False Accept* (FA), and the probability of rejecting genuine users, *False Reject* (FR). These parameters are defined in Section 5, along with other error rates like *Failure to Enroll* (FTE) and *Failure to Use* (FTU) rate.

The design of biometric authentication is an intricate process because of all the dependent parameters. When designing a biometric authentication application, the desired levels of the parameters are often specified, mostly in the form of maximum allowable FA and FR rates.

Storage requirements and throughput are quantitative parameters too, but these are parameters of the application or of the biometric and are discussed in Section 9 Section 10.

Qualitative parameters of biometric authentication and biometrics installations are the following three properties as discussed in Section 6 that go beyond the raw accuracy of the matching subsystem:

1. *Security:* This is the safety of the application achieved by eliminating all *points of attack* on, among other things, *the assets* of an application, i.e., protection against *interception* [152]. For a financial application, the asset is money. For a travel application, the actual transportation system (application) and the travelers are the asset. Hence the asset is people while the *threat* are people too. This has some implications that are discussed in Section 6.1.
2. *Convenience:* There are many aspects to convenience of a biometric application as discussed in Section 6.2. In essence however convenience is some measurement of how smoothly, efficiently and effectively the biometric authentication is implemented and functions while operating at acceptable error rates.

   In general, this is convenience for the enrolled population **M**. Convenience does not necessarily refer to convenience for the system operation.

3. *Privacy:* This, of course, is an authentication system property that is not easily defined. Privacy, of the enrolled population, is more specifically discussed in Section 6.3, but just like convenience, privacy, however defined, is a function of other biometric system parameters.

There has been much interest in the *individuality* of the various biometric identifiers. Loosely speaking, this has to do with comparing a biometric with a password [134]; and it is related to how easy it is to randomly "guess" a biometric machine representation, given current sensing capabilities.

Intrinsic error rates: This is yet another term that refers to this individuality of biometric identifiers. There exists the notion of a biometric's *intrinsic* error rates, somehow associated with the nature of the biometric. The speculations and claims about these intrinsic numbers can be astounding and it should be understood where these numbers might be coming from. (Error rates are estimated using techniques described in Section 11.)

There are some interesting legal cases on about biometric intrinsic strength. More on these intrinsic error rate of a biometrics can be found in Section 8 where the concept is explained. This will also bring us to the notion of 'weak' and 'strong' biometrics and allows comparisons of the strength of a biometric to passwords.

Security versus cost

In the final analysis, biometric authentication is about securing an application, at a certain cost:

The cost to install and maintain an application depends very much on the various choices and compromises that are made to achieve the desired levels of the above parameters and more qualitative application properties, such as, convenience and privacy.

Adding automated biometrics to a (legacy) authentication system surely will add to the operating cost of the application. This definitely is the case when the motivation is increased security and not cost savings. When the motivation for the biometric is increased convenience (no need to remember a password) it is not immediately clear that this will lead to cost savings.

Cost is naturally determined by system specifications in term of parameters and properties described in Sections 5 and 6.

Selecting a biometric

Given a biometric authentication application, the cost and security of the installation, no doubt, depend on the choice of biometric and selecting a specific biometric for an application is of prime concern. Figure 6 indicates that the error rates are of course a main decision factor. The properties of the biometric and of the application are intricately related, not just based on cost factors, but also because the application (e.g., the users) can have enormous impacts on accuracy. These type of issues are discussed in Sections 7-8.

There are other characteristics that differentiate the various biometrics and differentiate the applications:

Application properties: These are issues related to the application. For example, an application may need such high security that an unusual biometric, the retina, is used as biometric.

On the other hand, with telephone communications so pervasive and the cost of microphones so low, voice biometrics is the obvious choice for remote secure authentication applications.

Section 9 describes properties of biometric applications.

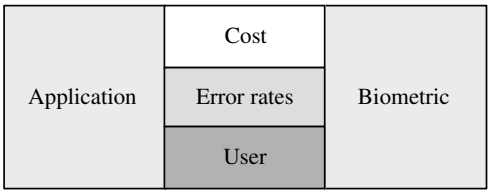


Figure 6: The choice of biometric is *not* just based on error rates.

Biometric properties: There are characteristics that can vary greatly from biometric to biometric, such as *intrusiveness* of the biometric acquisition and the social stigma surrounding a particular biometric.

Many of these properties are qualitative and described in Section 10.

2.3 Part III: Meta-issues and advanced topics

Part III (Sections 11-15) describes some topics that are of great interest to the research community, e.g., the issues around enrollment in, see Section 12. Other topics are *biometrics integration* (Section 14) and things like *standards* (Section 15). However, we feel that the reader can initially just read this particular section (Section 2.3), rather than Sections 11-15.

The advanced topics handled in Part III are:

Testing and training: An issue is, the probably still somewhat misunderstood notion, that biometrics authentication applications have to be *trained* using biometric data and therefore *tested* using biometric data (Section 11). The training data is typically acquired from volunteer subjects. Ideally the authentication system is trained on the users (friends) and simultaneously on the "real" enemies (foes) of the application. The issues around *friends* and *foes* of the application are discussed in Section 12.

System training and testing then is iterative optimization, as shown in Figure 7. For the verification database **M**, ideally a database M that is a good approximation to **M** should be selected. Similarly, a test set N should be defined that approximates **N**, the foes.

A biometrics authentication system deals with test databases M and N that are more or less a *segment* of the world population. The problem here, of course, is that it is very hard, close to impossible, to mathematically or algorithmically model the biometric of a subject as a function of population

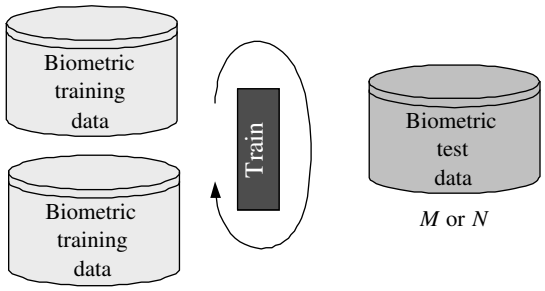


Figure 7: System design is a matter of training and testing to achieve the design specifications.

segment, i.e., the customer population of the application **M**. Therefore, the biometric authentication application has to be somehow *trained,* there is no way around it. How biometric authentication systems are optimally trained using biometrics data (especially during operation of the biometric authentication system) is still a wide-open topic of research.

Ultimately, this training will have to take place after a biometric application has been installed at the physical site and during the enrollment of subjects. Simultaneously both enrolling the population and training the authentication system obviously confounds the problems. This, and other issues involved with training, are discussed in Section 11.1.

During testing, the parameters of a biometric authentication application are statistically estimated. Here again databases of biometric samples acquired from sets of subjects are used. The biometric samples are properly labeled as to subject identity and hence estimates such as incorrect match rates can be computed. It should be noted here that the estimates of the errors are just estimates and can be made to look arbitrarily good or bad by simply choosing the test data sets accordingly.

The issues surrounding biometric system testing are perhaps the most prominent reason for much of the biometrics hype and confusion. Many different vendors of different biometric systems quote different numbers. Section 11.2 discusses testing procedures and explains how best to interpret numbers that are associated with biometrics or biometric applications.

Evaluation of a biometric authentication protocol and application does *not* measure the *intrinsic* strength (see Section 8) of the biometric itself per se. The parameters, such as error rates, can only be estimated when a biometric application is implemented in an end-to-end biometrics authentication. *The effectiveness of a particular biometrics can only be assessed in light of the application.*

Enrollment of subjects: *Enrollment* is the actual, possibly assisted by humans, registration of subjects (people) into some database by acquiring biometric samples and storing some representation of the samples. Hence, after enrollment, a subject is *just* some machine representation. By definition, the enrollment part of the application has many of the properties associated with biometric authentication as described throughout this document. Enrollment itself involves authentication and determining the eligibility and/or identity of a subject through some authentication protocol.

Issues specific to enrollment, such as *ground truth*, are discussed in Section 12. Biometric enrollment bears similarities to biometric authentication hence the enrollment of a biometric application can be expressed in quantitative terms to a certain extent. Problems with false enrollment, something that could easily happen, are often not directly related to the use of biometrics though.

Enrollment policies: Such policies are central to biometric authentication. Enrollment policies that determine who are the eligible subjects **M** and ineligible subjects **N** need to be defined for each application. This includes things like:

- Who is to be enrolled in the application; who are the customers?
- Ideally, one would need to define who is not to be enrolled in the application.
- The "true" identity of an enrolled subject has to be verified or established somehow. What proofs of identity are accepted at enroll time?
- What credentials are issued and what credentials are shared with the application (biometrics, personal knowledge) for subsequent authentication. As we will see in Section 3, choices here will limit the flexibility of the authentication protocols.

- At enrollment, other databases may be available and subjects could be screened against these possibly large databases. The details of the policies for running such searches may or may not be publicly made available. Nevertheless, the enrollment policy somehow has to define this.

Note that there are two types of enrollment procedures, voluntary procedures to enroll for certain privileges and involuntary procedures for criminal databases.

Security attacks: Biometric authentication for securing an application against unauthorized access has many of the same security issues in common with traditional authentication [118, 152]. Beyond that, additional points of attack are introduced due to the use of biometrics *itself* as described in Section 13.

When biometrics are introduced to guarantee additional security, the possibilities of hostile attacks on the integrity of biometric data have to be considered. Probably the scariest scenario is a pointed and successful attack on the biometric authentication system where biometric tampering is used to violate the application. This relates to impersonalization attempts with cut-off fingers, etc.

For access control applications, potential threats are, among others, the users (rather abusers) of the application. Points of attack in such applications are also enrolled subjects (computer representations of subjects), i.e., the data integrity of logical and physical subjects (to use security terminology). Privacy attacks as described in Section 13 are, just like in traditional authentication applications related to the issue of confidentiality [118, 152]. Biometrics causes additional privacy concerns because recordings of personal traits (or machine representations thereof) are collected and stored.

Integration: The use of multiple biometrics for authentication is very much part of the authentication protocol, consider for example, (*face* AND *voice*) versus (*face* OR *voice*). These are two different protocols for using the biometrics face and voice together for authentication, see Figure 8.

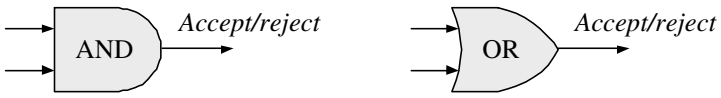


Figure 8: The integration method of multiple biometrics is part of the authentication protocol.

High error rates are common to some biometrics and often the use of multiple biometrics is therefore suggested. When properly combined, multiple biometrics lowers the probability of accepting intruders albeit at the cost of convenience.

Multiple biometric authentication can be assessed without any particular biometrics in mind, see Section 14.2. Integrating multiple biometrics is intimately related to authentication protocols and enrollment policies as discussed in Sections 3 and 12, respectively. The integration of multiple biometrics should be part of the secure authentication protocol.

Standards etcetera: Section 15, finally discusses the slow developments of biometric Application Programming Interfaces (APIs); the various standards that are being developed for, say, biometric data interchange; and the slow convergence on the agreements on testing and test databases.

3 Authentication and biometrics

Biometrics is about identification and verification of human beings' identities, or establishing trusted communication between two parties to negotiate access to an application. This is a topic that has been studied in the field of information security for quite a while now. Therefore, in the next sections we introduce some security terms and tools as can be found, for example, in [85, 118, 152]. Indeed we take the liberty to follow Chapter 4 of [118] fairly closely.

But first let us give some definitions pertaining to authentication that can be found in [85]:

- *Access control:* A mechanism for limiting use of some resource to authorized users.
- *Access control list:* A data structure associated with a resource that specifies the authorized users.
- *Authenticate:* To determine that something is genuine; to reliably determine the identity of a communicating party.
- *Authentication:* The process of reliably determining the identity of a communicating party.
- *Authorization:* Permission to access a resource.

3.1 Secure authentication protocols

Using conventions between two parties, cryptosystems can be used for purposes other than just secret communication, these conventions are called protocols. A *protocol* is an orderly sequence of steps two or more parties take to accomplish some task. The order of the steps is important in this activity, so the protocol regulates behavior of both parties. The parties agree to the protocol or at least the protocol is understood.

Let us use a telephone conversation as an example. Upon dialing, the dialing person hears a telephone ringing at the other end and, after a while, a click when the phone is answered by the second party. Protocol, or standard practice, is then that the receiver speaks first saying "hello" or even gives some identifier like a name. The originator then identifies him or herself. It is only after this protocol that the intended communication is initiated.

By simply picking up the phone and not answering, the reader can easily verify the utter failure in communication if this established protocol is not followed. Even if the dialing party hears the click, without this confirmation of connection, the dialing party will most often not initiate communication.

The initiation of a phone conversation is an example protocol. An authentication protocol should have the following desirable characteristics [118]:

- *Established in advance*: The protocol is completely defined and designed before it is used. The work flow of the protocol is defined, the rules that determine the work flow are defined, and it is defined what it means that two authentication credentials match.
- *Mutually agreed*: All parties to the protocol agree to follow the steps, and agree to follow these steps in the prescribed order.
- *Unambiguous*: No party can fail to follow a step properly because the party has misunderstood the step.
- *Complete*: For every situation that can occur there is an a priori defined action to be taken. This means, for example, that the exception handling process is completely defined.

The modern heavily networked and traveled world requires the use of computers and communications as tools for access to services, privileges, and applications. Users of a system are typically not acquainted with managers and other users of a system and communications are over long distances. Because of anonymity and distance, a user will not and should not trust the managers or other users of a system. Protocols need to be developed by which two suspicious parties can interact, these protocols in essence regulate behavior. Authentication then becomes the following of a behavioral protocol between user and application so the user can be authorized to use the application or to access the premises.

In itself, a protocol does not guarantee security. For example, the protocol that controls access to a church may specify the opening and closing hours and may require a certain dress code but does not contribute to the safety of the church.

3.2 Access control security services

Two security services, at least, should be offered by any access control system [152]:

- *Authentication:* This service is concerned with assuring that a communication is authentic. In the case of ongoing communication, two things are involved: (i) At the time of connection initiation, the service ensures that the two entities are authentic, i.e., that each entity is the entity it claims to be. (ii) The service must assure that the connection is not interfered with by a third party masquerading as one of the legitimate parties.

- *Non-repudiation:* This prevents either the sender or receiver from denying a transmitted message. Thus, when a message is sent, the receiver can prove that the alleged sender in fact sent the message. Similarly, when a message is received, the sender can prove that the alleged receiver in fact received the messages.

  This is perhaps best called reciprocal non-repudiation, which is mostly not incorporated into authentication systems, between a human subject and a physical or logical system.

So, from a security point of view, authentication is *only one* of the security services that access control should offer; there are a few more required services, like *confidentiality*, see [152]. An authentication system is then organized as outlined in Figure 9. A user interface A gathers credentials from a subject through input devices such as smart card readers and fingerprint scanners; the user interface may also include output devices, possibly to give feedback on the quality of the biometric sample that has been acquired. The access control system B offers a number of security services, among which authentication. The authentication is ensured through a prescribed protocol. Upon successful fulfillment of this protocol, the subject is given access to the application C, through the mechanical or logical operation of a switch d.

The access controls system, as indicated in Figure 9, also offers traditional security services such as confidentiality and non-repudiation.

In the following we look at some authentication protocols as they have evolved over the centuries and more recently have been established due to developments in photography and computing machinery.

3.3 Authentication methods

There are three basic modes for authorizing subjects, described by Miller in [106] and shown in Figure 10. These have evolved over the ages with the technology of the printing press, photography, and automation; the methods have been in use long before the widespread needs for automated, electronic authentication:

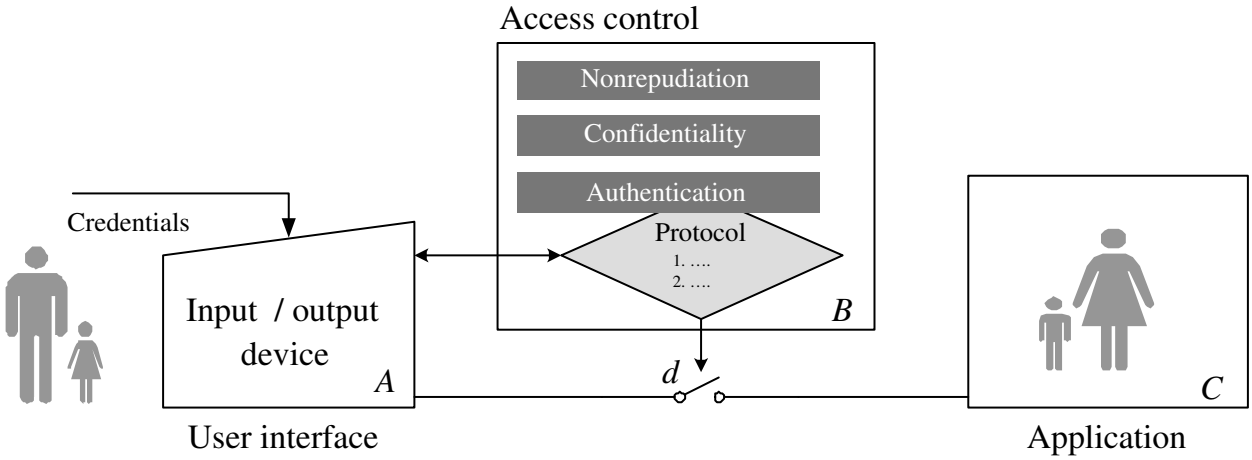


Figure 9: Security of access control involves authentication of both user and application.

P *Possession:* Anyone in the possession of a certain article, e.g., keys or card key, is eligible to receive the associated service or access. For example, anyone with access to the car keys has the privilege to drive the car.

In this document, possession is defined a little more broadly to include things like User IDs and account numbers. These are more public descriptors or identifiers of computer, bank, and other type of accounts and typically these numbers or descriptors represent assets and virtual assets.

K *Knowledge:* Individuals with certain knowledge are eligible to access the service. Authentication here is based on secret knowledge, such as, passwords, lock combinations and answers to questions. The important word here is *secret;* the knowledge ideally needs to be secret in order to be used for authentication.

However, well established knowledge tokens also include things like *"mother's maiden name"* that are likely to be known by a genuine subject but not by an intruder.

Knowledge can moreover include other personal facts about a subject that probably are only known to this subject. Examples of these types of knowledge are, the easily remembered little facts like: *favorite color, children's names,* et cetera. These facts can be used in dynamic authentication protocols as described in Section 14.3.

(This type of knowledge is not a biometric *mainly because it can be shared* some particular knowledge types also can be *changed at any time*.)

B *Biometrics:* Personal traits of humans that can be somehow measured (sampled, acquired) from a person in the form of a biometric identifier and that more or less uniquely distinguish a person with respect to the rest of the world population. These are properties that are somehow intrinsically related to a human and are largely determined by either genetics or phenetics, inheritably characteristics as opposed to characteristics formed during the foetus phase. They are difficult to share, steal, or forge.

There is a fine line between authentication mode P, possession, and authentication mode K, knowledge. Consider the following. If one is in the possession of a phone, one is in the possession of a phone number. In

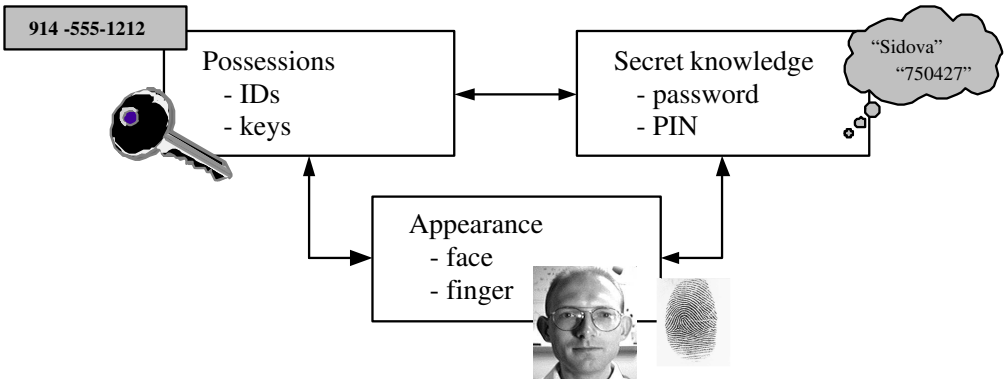


Figure 10: The three basic ways a person can prove identity.

a sense, the telephone number is knowledge, but it is mostly not secret and it is at least known to a group of people. The phone number is therefore more of a possession type of authentication mode because it uniquely identifies its owner, just like secret knowledge will uniquely authenticate the person with the knowledge.

One biometric, namely signature, is more related to knowledge and writing skills to put a signature on paper. Hence, signature is a biometric that has knowledge characteristics.

There are other confusions between knowledge K and biometrics B as will be discussed later.

Table 3 shows four methods for user authentication that are widely used [106]. Because a biometric is an intrinsic property of some individual, they are difficult to surreptitiously duplicate and nearly impossible to share; additionally, a biometric of an individual can be lost only in the case of serious accidents, severe illnesses, and extreme wear and tear on the biometric. Therefore, biometric identifiers offer certain assurances about the real identity of the user in an authentication protocol; something that the use of other modes for authentication, possession and knowledge, do not guarantee. When combining the last row (B) with possession P and/or knowledge K, we get additional biometric methods like (P, B) (e.g., passports, smart cards with biometric template) and

$$(P, K, B) = (P = credit\ card, K = \text{“}mothers\ maiden\ name,\text{”}\ B = signature),$$

a much used authentication method for credit cards.

Of course, biometrics have many drawbacks too, for example, a face appearance changes because of facial hair and haircuts, a voiceprint may be influenced by illnesses like the flu and colds, etc. This gives rise to many research issues as described in the rest of this document.

3.4 Possession and knowledge

Possession (as used in this document) and knowledge, in the form of $(account\ number, password) = Possession, Knowledge) = (P, K)$ is probably the most widely used authentication method (protocol). This authentication method is used for computer accounts, Internet accounts, intranet accounts, email account, phone mail, etc. These authentication modes P and K only require exact matching but do not link a user (a real person) to some more or less established "identity." But rather they link some *well-defined* identity determined by possession to the *anonymous knowledge of a password* and *not* to the authentic enrolled person.

Method	*Examples*	*Properties*
What you have (P)	User IDs, accounts Cards, badges Keys	Can be shared Can be duplicated May be Lost or stolen
What you know (K)	Password, PIN Mother's maiden name Personal knowledge	Many passwords are easy to guess Can be shared May be forgotten
What you have and what you know (P, K)	User ID + Password ATM card + PIN	Can be shared PIN is a weak link (Writing the PIN on the card)
Something unique about the user (B)	Fingerprint Face Iris Voice print	Not possible to share Repudiation unlikely Forging is difficult Cannot be lost or stolen

Table 3: Primary existing user authentication methods with example positive/negative properties.

Authentication mode B (biometrics) offers additional security because it is hard, to impossible, to share biometrics and therefore the one security aspect of access control, namely authenticity of the participants, is much more confidently and therefore securely established.

3.5 Authentication protocols

As pointed out in [17], any authentication protocol with multiple modes (*and* multiple biometrics) can be defined and can be executed against a set of presented credentials. An *authentication protocol* is the (automated) decision process and work flow to determine if a subject's credentials are sufficient proof of identity to authorize the subject for access based on credentials, or tokens.

The first column of Table 3, are the necessary credentials that a user needs to have in order to be authenticated. Here the term "authenticate" is used in the strict security sense of the word, "the process of reliably determining the identity of a communicating party." In order of the rows of the table we have:

1. This can be the possession P like physical key to gain access to a certain locale. This can be the simple possession of a credit card at a gas station that enables the owner to authenticate credit card transactions (at least in the US).

2. In the case of a vault, the knowledge K is the combination to the lock that authorizes a person to enter the vault; this key knowledge implicitly authenticates removal of objects from the fault. Here the authentication protocol is simple, a user identifies him, herself to the vault through the secret knowledge of the pass code.

3. A well-known everyday authentication protocol requires an ATM card P and a PIN K for banking through an automated teller machine. Anyone that has these credentials *and* is aware of the rest of the authentication protocol, is authorized to access a banking application and therefore can perform transactions through secure authentication.

4. The fourth row in Table 3 refers to the pure identification protocol, which is simply the presentation of a biometric *and nothing else*, no other tokens or forms of interaction with the user interface of an authentication system are involved.

   Without the concept of biometrics, i.e., invariants over time in human appearance, there is no hope of solving the automated human identification by machines processing sensory signals. It is biometric identifiers that truly distinguish one *person* from the next.

Hence for any secure authentication protocol there are one or more authentication modes involved that we call *tokens,* hence we have possession tokens P, knowledge tokens K, and biometric tokens B. The enrollment then (as discussed in Section 12) is a communication between user and the access control system to exchange authentication tokens. The system supplies possession and knowledge tokens to the user, while the user may supply biometric tokens (samples) to the system. This process is defined by an *enrollment policy* (Section 12).

A set of tokens $T = (P, K, B)$ is only part of an authentication protocol. What is needed further is a set of rules R that define the authentication protocol that uses $T = (P, K, B)$ according to precisely defined orderly sequences of steps, or rules of behavior, as defined above, denoted as

$$R = R(T) = R(P, K, B).$$

Combining multiple authentication methods, especially biometrics B, into an *authentication protocol* improves the certainty of authentication and decreases therefore chances of repudiation and fraud. For example, an authentication protocol as

$$R = R(P, K, B) = (P, K, \{B_1, B_2\}) = R(credit\ card, PIN, \{photograph, signature\}) \qquad (1)$$

specifies what authentication modes are used in a protocol and specifies how these modes are to be used in the protocol, R, a set of rules operating on $(P, K, \{B_1, B_2\})$. The actual authentication protocol R in (1) may be described as, loosely: Anyone in the possession of a credit card P with a signature and a picture and has the ability to produce a signature B_1 that appears similar to the signature on the credit card and has a likeness B_2 to the picture, and additionally has knowledge K of the associated PIN, has the privilege to use the credit card.

In fact, not surprisingly, authentication protocols are found throughout the biometrics literature. A few examples are:

- The authentication rule: "*three tries and you're out*." A subject is given three chances to match the reference biometric but after the third failure, the subject is denied further access to the system.

- The Galton-Henry system [6] of manual fingerprint classification using ten-print cards, which was published in June 1900 and officially introduced at Scotland Yard in 1901 for its criminal-identification records is an authentication protocol. Fingerprints are classified in a three-way process: by the shapes and contours of individual patterns, by noting the finger positions of the pattern types, and by relative size, determined by counting the ridges in loops and by tracing the ridges in whorls.

  Adaptation of the Henry system for computerized large-scale searches [109] is an automated authentication protocol.

- The fascinating science of latent fingerprint identification is described in very precise minutiae matching protocols [116], thereby defining as precisely as possible *what it is* that a latent print matches with a reference print.

Method	*Examples*	*Properties*
Something you have with something unique about you (P, B)	Passport with face image Credit card with signature	Forging is easy Easy to share
	Smart card with biometric	Hard to tamper with

Table 4: Some existing user authentication methods that use biometrics.

- Manual identification for law enforcement purposes follows a protocol. Here all the authentication modes and methods are used in an interrogative and investigative way, i.e., some long string of user tokens $P_1, B_1, B_2, K_1, P_2, B_3, ..., X_m$ with biometrics modes $B_1, B_2, B_3, ...$ is matched to tokens in a multitude of (government agency) databases.

  This of course is the optimal way of positively identifying a subject. Clearly the frequent use of multiple biometrics in law enforcement greatly enhances the accuracy in terms of error rates of the investigative process thereby decreasing false conviction rate. From the point of view of biometrics identification, the false positive rate is minimized.

Hence, in addition to a set of tokens $T = (P, K, B)$ and rules operating on these tokens $R(T)$, an authentication protocol needs to define what it means that two tokens of any kind, P, K, and B "match."

3.6 Human verification

When people were still living in small communities there was really no need for authentication methods that explicitly require a specific biometric. A possession authentication method such as a name has sufficed long as a primary verification method. Biometric methods such as face (and voice) were in use of course but before photography there was no possession authentication method with a face image. What was in use however was some loosely organized database of (name, face) pairs in each person's head, in addition to some screening database of faces of undesirable individuals. The travel and communication industries rapidly increased the number of faces (and names) that someone has to know. The human brain, remarkable as it is, has never been designed to solve large scale face image matching or searching. Therefore in addition to the authentication methods in Table 3, authentication method (P, B) are in widespread use.

Due to many convenience factors and its naturalness, "face" has evolved as biometrics of choice for manual authentication with probably as second runner-up maybe "signature." These are the examples in the first two rows of Table 4, though signature is almost never manually verified when using a credit card (in the US). The direct extension of this (P, B) authentication method is the smart card with a stored biometric. In terms of acceptance, people have used possession plus a biometric as an authentication method for many years. Of course, it is very hard to estimate at what false accept and false reject rates human beings can verify faces.

Clearly, there is large variation in verification accuracy from one person to the next and verification accuracy depends on the state of fatigue of the human inspector. There is however no reason to believe that a human can perform face verification very well, the error rates are in the order of 1 in 1000, or 10^{-3}.

The error rates for human verification of signatures are hard to establish. It is well known that a handwriting experts can do signature verification very accurately, but clearly the cursory inspection of signatures on credit card slips is not much of secure process.

More on authentication modes and protocols can be found in [17].

In Section 3.5, we pointed out that for any authentication protocol there exists a set, or triplet, of tokens $T = (P, K, B)$ that are the required credentials for access to an application. Here P are possessions, K is knowledge, and B are one or more biometrics. For authentication, the credentials of a subject need to be matched, including the biometric templates. For biometric user verification, this amounts to an authentication protocol that is the same as for today's password systems, see Figure 11. A subject lays claim on an identity, or rather a numeric or symbolic identifier like an account, but instead of exact matching of a password, the authentication protocol of a biometric system requires matching one or more biometric samples, which can only be done probabilistically or in some fuzzy fashion.

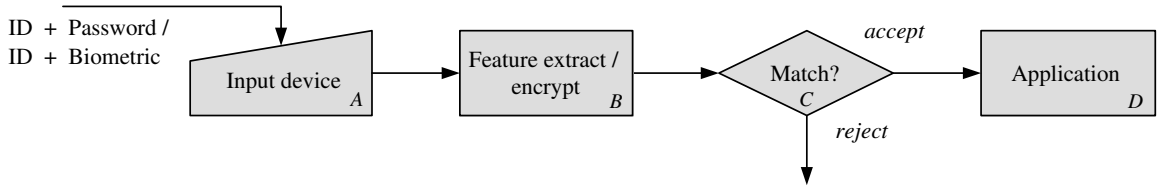


Figure 11: In essence biometric authentication is not that different from a password authentication.

In terms of the authentication system Figure 11, A is the input device, a keyboard or some sensor; B is the password encryptor or in case of biometric input a feature extractor; C is a matcher that matches either passwords or biometric templates; and, D is the application protected by the access control system.

The introduction of biometric authentication protocols brings with it the concept of inexact matching, and inexact matching of multiple biometrics. The latter is of course clearly a part of the authentication protocol. We discuss these issues and other things in this section.

3.7 Matching and fuzzy matching

The ability to match, both parametric identifiers, such as passwords and knowledge, and biometric identifiers is at the heart of the problem of biometric authentication.

The decision to authenticate is based on one or more authentication methods, or tokens from the triplet $T = (P, K, B)$. For person authentication, each token supplied by a user needs to be matched with the token as it is recorded and stored during enrollment. The decision on whether these tokens match or not is to be made by integrating the output of the separate match engines that verify the tokens or by integrating the token matching more fundamentally. Comparing possession tokens and simple knowledge tokens such as passwords amounts to exact matching, though it is conceivable that for low security transactions, passwords that do not match exactly can be accepted. For example, a PIN code may be accepted if two numbers in the PIN that is entered are transposed.

Two issues here need attention:

1. *Integration of credentials:* It is best of course to combine two or more authentication methods. As noted before, associating possession P (an ID) or knowledge K with a biometrics B reduces the problem of biometric identification to biometric verification, i.e., reduces the problem from the $1 : m$ matching to $1 : 1$ matching of biometrics.

2. *Integration of biometrics:* The required credentials T may include multiple biometrics, i.e., $(B) =$

$(\{B_1, B_2\}))$, with B_1 say finger and B_2 face. Integrating multiple biometrics is and has been a subject of considerable interest and will be discussed later.

Integration of different biometrics is one problem, clearly a problem that lies in the domain of pattern recognition. Integration of different non-biometric authentication modes, on the other hand, is very much a topic that is part of traditional security of authentication applications (e.g., [152]). Integrating traditional authentication methods with biometrics is only beginning to be studied together in the context of the overall security architecture of the application [17, 131].

Again, in general, the use of any of the above modes P, K, or B means that one has to be able to match, such as manually or machine verification of account numbers or pass phrases, and manual comparison of appearances, such as faces. Possession and knowledge tokens require pretty much just exact matching if done by machine. Biometric matching, on the other hand, involves some type of fuzzy matching as we will explain in the rest of this document.

Hence, the fundamental difference between biometric identifiers (tokens) and other authentication methods (tokens) is the notion of the *degree of a match*, i.e., the underlying matching technology. A password authentication protocol always provides a crisp result: if the passwords match, it grants access and otherwise it refuses access. That is, there is *no concept of the probability of a match*. Consequently, there is no confusion about the *precise definition* of a match. By necessity, however, biometrics has to adopt probability theory and the use of statistical techniques. This has resulted in the concepts of *application error rates* (false accept and reject rates) and *intrinsic error probabilities* (loosely the minimal achievable error rate for a given biometric) that are associated with biometric authentication systems and biometric identifiers. These are discussed in Section 5 and Section 8, respectively.

Poor definition of *what it is exactly* that is estimated for a particular biometric installation or biometric and controversial, or dubious, testing methods contribute much to the hype surrounding biometrics as we will see later (Section 7).

3.8 Screening

Screening is negative identification, establishing that a person is *not* on some watch list. A research area that is developing itself is passive "screening" of crowds based on face images in a crowd and face cataloguing. Face cataloguing is a concept that is being popularized by the Visionics Corporation [161] and is the technology of building a repertoire of face images of people in a space, based on ordinary visual light cameras.

In general, screening is the authentication protocol that prescribes matching all tokens, or credentials, of a subject (passenger) to a variety of government and civilian databases. The authentication protocol defines that a subject can be authenticated if a string of the subject's tokens (credentials) $P_1, B_1, B_2, K_1, P_2, B_3, ..., X_m$ does *not* match with "credentials" of the list of most wanted criminals. Biometrics can play a role in this screening process; however, with the current state of the art none of the biometrics can be expected to narrowly pinpoint down true subjects based on a biometric alone (see Section 7.2). Such biometric matches will create many false positives, too many to handle if the list of the criminals in the database is large. Biometric identifiers can be used on the other hand to match against the return lists of parametric searches, based on name, date of birth, etc. Such lists may be sufficiently small to be able to make decisions based on biometric matches.

As already explained in Section 3.5, manual positive identification of an (arrested) subject for law enforcement purposes follows a protocol. Here all the authentication modes and methods are used in an interrogative and investigative way, i.e., some long string of user tokens $P_1, B_1, B_2, K_1, P_2, B_3, ..., X_m$ with

biometrics modes $B_1, B_2, B_3, ...$ is matched to tokens in a multitude of (government agency) databases. The tokens $P_1, B_1, B_2, K_1, P_2, B_3, ..., X_m$ are the possessions of the suspect, knowledge tokens extracted from the suspect and biometric samples measured from the subject. When a match occurs this is indication of a possible identity of the subject.

This of course is the optimal way of identifying a subject. Clearly the frequent use of multiple biometrics in law enforcement greatly enhances the accuracy in terms of error rates of the investigative process thereby decreasing false conviction rate. From the point of view of biometrics identification, the false positive rate is minimized.

3.9 Continuity of identity

A biometric might not be matched just once at some access point, but could instead be continually verified while a person is in some restricted physical space or accessing other resources. This ensures what is called "continuity of identity." For example, for voice and face biometrics, verification can take place continuously or periodically in the background as needed (when fraud is suspected in the middle of the transaction, for instance), or at anytime after the transaction is completed by recording the biometric signals and analyzing it later for a match. Verification can also take place in an incremental manner, and the user may be granted higher privileges if higher verification probabilities are obtained with more biometric data collected as the interaction progresses. The verification data can also be used to update the biometric templates online, thereby learning or refining the machine representation of the biometric. Parallels to this incremental, continual authentication can be identified for almost any biometric.

The next line of research, related to surveillance just like face cataloguing is the issue of tracking people through a space (see the proceedings of a recent workshop in tracking [53]). What the ultimate goal of this research is the *continuity of identity* aspect of the surveillance problem. After all, if a person is authenticated at a point A, there is no guarantee that at point B this person (with credentials, like boarding pass) is still the same identity.

4 The most common biometrics

The most commonly used automated biometric identifiers are: (i) finger; (ii) face; (iii) voice; (iv) hand geometry; (v) speaker (voice) and (vi) signature. Retina recognition or "eye recognition" is often mentioned as a possibility but most everyone means iris when mentioning "eye recognition."

Retina is used in high security prison applications [62]. Retinal authentication systems have harder-to-use input acquisition devices, since is has proven difficult to image the back of the inside of the eyeball.

We provide a brief description of the most widely used (or widely discussed) biometrics.

4.1 Fingerprint

Fingerprint has by far the longest and most interesting history of any biometric [92]. The inside surfaces of hands and feet of humans (and, in fact, all primates) contain minute ridges of skin, with furrows between each ridge, see Figure 13. The purpose of this skin structure is: (i) Exudation of perspiration, (ii) Tactile facility, and (iii) Provisions of a gripping surface. Just like irises, fingerprints are of the phenotypic type and hence fingerprints are not determined by genetics. In [117], it is shown, for instance, that identical twins have fingerprint that are quite different. Fingerprints indeed are distinctive to a person, in fact: "no one has ever been found who has a sequence of ridge detail on the hands and feet that is identical to the ridge detail of any other person." There is evidence that, e.g., the Chinese were aware of the individuality of fingerprints well over 5,000 years ago [6]. All this led to the widespread use of fingerprints in law-enforcement identification applications.

Early in the 20th century, an ingenious recognition system based on ten prints developed by Sir Edward Henry was brought in operation [6]. This system is now known as the "Henry System" and adopted and refined by the FBI [51]. It allows for correct identification of offenders by manual indexing into databases of known criminals. It classifies the overall flow pattern of the fingerprint into a number of distinct patters such as "arch", "left whorl", "tented arch", etc. These classes are not uniformly distributed over the population [6, 51]. All ten fingers are then classified in this way to yield a signature vector of the form $[A, W_L, W_R, A_T, L, ...]$ (usually referred to as "ten-print cards"). While not unique to each person, this sequence can at least be used to rule out *some* suspects. Much research has been done on this sort of automated fingerprint classification, e.g., [27, 73, 145], for AFIS applications.

In the early sixties, however, the number of searches that needed to be done on a daily basis simply became too large for manual indexing and Automated Fingerprint Identification Systems (AFIS) started to be developed [109]. Just like the matching of *latent* prints that have been found at crime scenes [51], an AFIS still requires much manual labor because a search may result in many false positives. In this document, we concentrate of on fingerprint matching for authentication ($1:1$ matching) but, of course, we keep fingerprint matching for screening applications in mind.

4.1.1 Fingerprint matching

Figure 13 shows the problems in fingerprint matching very clearly. A fingerprint authentication system reports some degree of similarity or some sort of "distance" (dissimilarity) between two fingerprint images and should report these measures accurately and reliably, irrespective of all the imaging problems discussed in the next section. Ideally, the similarity between two impressions as in this figure should be large, or equivalently the distance between the images should be small. Hence, the similarity or distance between two impressions of the same finger should be invariant to (i) translation, (ii) rotation, and (iii) elastic distortion between the impressions due to the elasticity of the finger skin.

Fingerprint matching has been studied over several decades by many researchers [107, 108, 127, 170]. Two broad classes of matching techniques can be distinguished, i.e., techniques that match images and techniques that match features:

- *Image techniques* (e.g., [30, 59, 78, 94, 126, 137, 153]): This class includes both optical as well as numerical image correlation techniques. Several image transform techniques have been also explored. These matching techniques will become important when the finger surface area that is sensed is small as in, e.g., CMOS sensors (see below).
- *Feature techniques*: This class of techniques extracts interesting landmarks (features) and develops different machine representations of a fingerprint from these features. This is the most widely used approach to fingerprint matching, which we will discuss in more detail.

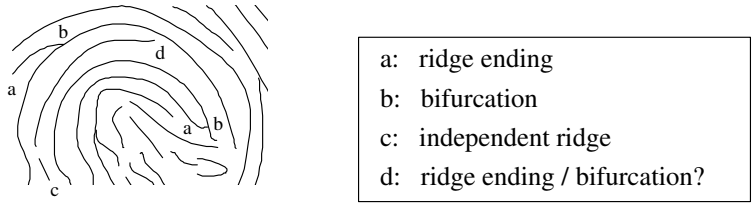


Figure 12: Ridge patterns of individual fingers have minute detail that distinguishes one print from another.

Beyond fingerprint class, the human expert uses many minute types of features of the ridge flow pattern to determine if two impressions are from the same finger. Figure 12 shows a piece of thinned fingerprint structure with a few examples of these features: (a) ridge endings, (b) ridge bifurcations, and (c) an independent ridge ([6, 51] give more complete listings, such as, lake, spur, crossover). Automated fingerprint matching algorithms attempt to match fingerprint impressions in a similar fashion. However, the most commonly used fingerprint features are only *ridge bifurcations* and *ridge endings*, collectively known as *minutiae*, which are extracted from the digitized print. Many matching algorithms do not even distinguish between bifurcations and endings because during acquisition and fingerprint image processing, depending on the amount of pressure exerted by a subject, a ridge ending may change in a bifurcation and *vice versa*. For example, for feature (d) in Figure 12 it is unclear if it is an ending or a bifurcation.

A thinned image as in Figure 12 is a step in the process of fingerprint feature extraction. This process typically starts by examining the quality of the input image as discussed in Section 12.4 (see also, e.g., [177]). Virtually every published method of feature extraction then (e.g., [101, 130]) proceeds by computing orientation of the flow of the ridges, which reflects the local ridge direction at each pixel. The local ridge orientation is then used to tune filter parameters for enhancement and ridge segmentation. From the segmented ridges, a thinned image (Figure 12) is computed to locate the minutiae features. Usually, a minutiae post processing stage cleans up several spurious minutiae resulting from either fingerprint imperfections (dirt, cuts), enhancement, ridge segmentation or thinning artifacts.

The machine representation of a fingerprint is, as noted before, critical to the success of the matching algorithm. A minimal representation of a processed fingerprint is a set $\{(x_i, y_i, \theta_i)\}$ of k minutiae, i.e., a set of points (X, Y) expressed in some coordinate system with a ridge direction at this point Θ as in Table 5. Such a representation and point set matching is used in [132]. The representation used by Jain et al. [75] is a string and matching is performed through string matching algorithms. Both these techniques do not take

into account the local topological information available in the fingerprint image. Graphs have been used in fingerprint analysis, primarily for fingerprint classification and matching. Isenor and Zaky in [68] use a graph representation, where nodes correspond to a ridge and edges to neighboring ridges or intersecting ridges. A graph matcher where nodes correspond to minutiae is presented in [129]. Other feature matching methods can be found in [149, 154].

X	Y	Θ
x_1	y_1	θ_1
x_2	y_2	θ_2
$\vdots$	$\vdots$	$\vdots$

Table 5: A minimal machine representation of a fingerprint image.

There exists a third class of algorithms for matching fingerprint images that combine the above approaches:

- *Hybrid techniques:* A third class of matching techniques [35, 76, 162, 174] combines both transform and feature techniques or uses neural networks in interesting ways to improve accuracy are considered. For example, Hamamoto [60] describes an identification method based on Gabor filters. Jain et al. [76] present a matching algorithm that uses features such as responses to Gabor filters of different frequency and orientation.

In general, a representation may be derived at the client end of the application or, alternatively, the raw image may be transmitted to the server for processing. Such transmission (and storage) of fingerprint images typically involves compression and decompression of the image. Standard compression techniques often remove the high frequency areas around the minutiae features. Therefore, a fingerprint compression scheme called as Wavelet Scalar Quantization (WSQ) is recommended by the FBI [23, 52].

4.1.2 Fingerprint image acquisition

In the 1980s, the development of cheap document scanning technology and personal computers enabled the use of fingerprint matching technology in everyday applications. Subsequently, the advent of several ink-less fingerprint scanning technologies coupled with the exponential increase in processor performance has taken fingerprint recognition beyond criminal identification applications to several non criminal, civilian applications such as access control; time and attendance; and computer user login. Over the last decade, many novel techniques have been developed to acquire fingerprints without the use of ink. These scanners are known as "livescan" fingerprint scanners. The basic principle of these ink-less methods is to sense the ridges on a finger, which are in contact with the surface of the scanner. The livescan image acquisition systems are based on four technologies:

- *Frustrated total internal reflection (FTIR) and other optical methods* [55]: This technology is by far the oldest livescan method. A camera acquires the reflected signal from the prism as the subject touches a side of the prism. The typical image acquisition surface of 1" $\times$ 1" is converted to 500 dpi images using a CCD or CMOS camera.

  Many variations of this principle, such as the use of holographic elements [105], are also available.

An issue with these reflection technologies is that the reflected light is a function of skin characteristics. If the skin is wet or dry, the fingerprint impression can be "saturated" or weak, respectively, and hard to process. These problems can be overcome to some extent by illuminating the prism with ultrasonic energy instead of visible light.

- *CMOS capacitance* [80]: The ridges and valleys of a finger create different charge accumulations when the finger touches a CMOS chip grid. With suitable electronics, the charge is converted to an intensity value of a pixel. Normally at 500 dpi these scanners provide about 0.5" $\times$ 0.5" of fingerprint surface scan area.

  This can be a problem as two impressions of the same finger acquired at two different times may have little overlap (see fingerprint matching techniques described below). The images also tend to be affected by the skin dryness and wetness. In addition, these CMOS devices are sensitive to electrostatic discharge (static electricity shocks).

- *Thermal sensing* [99]: This sensor is fabricated using pyroelectric material, which measures temperature changes due to the ridge-valley structure as the finger is swiped over the scanner and produces an image. This works because wet skin is a better thermal conductor than air and thus the ridges cause noticeable temperature drops on a heated surface. The technology is claimed to overcome the dry and wet skin issues of optical scanners and can sustain higher static discharge. The resultant images however are not rich in gray values, i.e., dynamic range.

- *Ultrasound sensing* [7]: An ultrasonic beam is scanned across the finger surface to measure the depth of the valleys from the reflected signal. This can theoretically be implemented as a non-contact sensor. Skin conditions such as dry, wet and oil on the skin do not affect the imaging and the images better reflect the actual ridge topography. However, these units still tend to be very bulky and require larger scanning time than the optical scanners.

Figure 13: Impressions of the same finger can be quite different due to elastic distortion.

Fingerprinting for person identification had an advantage over most other biometrics in that fingerprint acquisition has been possible for centuries in the form of impressions of inked fingers on paper but also the impression of fingers in materials like clay. However, there is a property of automatic fingerprint recognition systems that is not shared by many other pattern recognition systems, i.e., the process of sensing, or acquiring, the biometric involves *touching* some input device with the pattern *itself*. Because of this touch sensing, the actual pattern that is being sensed is distorted during the acquisition of the pattern. Figure 13 shows that this type of elastic distortion can be quite different for prints of the same finger – simply overlaying the images will not work. Recently, non-contact [40] fingerprint scanners have been announced that avoid problems related to touch sensing methods, including the elastic distortion of the skin pattern.

4.2 Face recognition

Face recognition is an instance of object recognition from images, a central problem in the area of computer vision (e.g., [3]). Recognition of objects from images of these objects is a fundamental problem and basically an unsolved problem. Probably the most important issues in the overall problem of object recognition from images are the problems of *segmentation* and *representation.* Segmentation refers to the partitioning of the input image into the object image, those pixels that arise from the object surfaces, and the background image. Representations are internal models of the 3D objects that one desires to automatically recognize in input images. Represenations are application-specific and attempt to preserve important properties but provide invariance or at least resistance to "noise" in the signal. In the case of nonrigid objects like faces, approaches to machine representations of human faces are often based on machine learning.

Face appearance is a particularly compelling biometric because it is one used every day by nearly everyone on earth as the visual means for recognizing other humans (Section 3.6). Since the advent of photography it has been institutionalized as a guarantor of identity in passports and identity cards (see Section 3.3). Because conventional optical imaging devices easily capture faces, there are large legacy databases (police mug-shots and television footage, for instance) that can be automatically searched. Because of its naturalness, face recognition is more acceptable than most biometrics, and the fact that cameras can acquire the biometric passively means that it can be very easy to use. Indeed, surveillance systems rely on capturing the face image without the cooperation of the person being imaged and can be done overtly and covertly, the latter of course raises privacy concerns.

Automatic face recognition has a 30-year history, starting with the face recognition system designed by Kanade [82]. Turk and Pentland [157] popularized face recognition by using face image transforms, which were introduced by Kirby and Sirovich [148, 87] for representing and compressing face images. Kirby and Sirovich also developed a computationally efficient matrix computation of the transform, this development and the publication of [157] resulted in a flurry of activities in face recognition research [29, 142].

This transform, originally known as the Karhunen-Loève transform [84, 96], is now known under names like Principle Component Analysis (PCA), etc.

In general, face recognition systems proceed by detecting the face in the scene, thus estimating and normalizing for translation, scale and in-plane rotation. Many approaches to finding faces in images and video have been developed; all based on weak models of the human face that model face shape in terms of facial texture and face appearance in terms of flesh tones. Approaches then divide [25] into appearance and geometric approaches, analyzing the appearance of the face and the distances between features respectively.

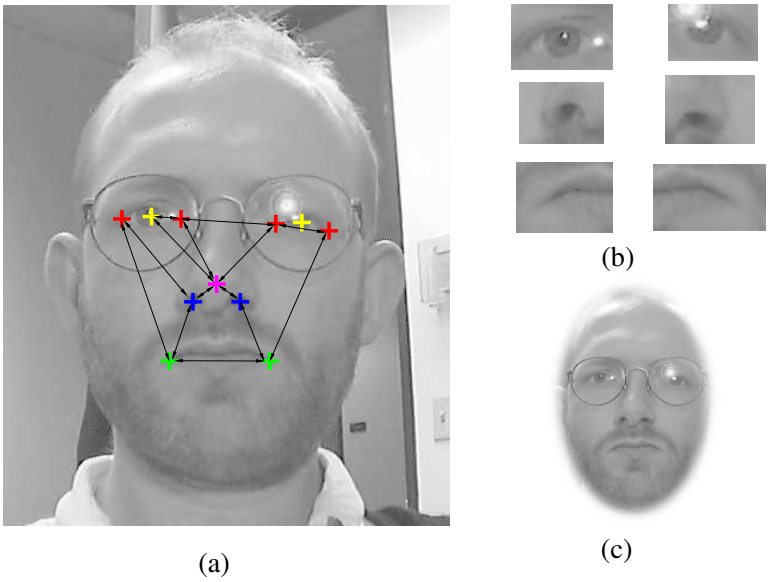
- *Face geometry approaches:* Here, of course, the idea is to model a human face in terms of particular face features, such as eyes, mouth, etc., and the geometry of the layout of these features. The work

from von der Malsburg (e.g., [175]) is perhaps the most sophisticated geometrical approach, where a face images is represented by a graph.

The goal of geometric approaches is to model a human face in terms of simple features and their relationships and extract these features from face images. Face recognition is then a matter of matching feature constellations.

- *Face appearance approaches:* The underlying idea behind these approaches is quite simple. A face image is transformed into a space that is spanned by basis image functions, just like a Fourier transform projects an image onto basis images of the fundamental frequencies. In it simplest form, the basis functions, Eigenfaces, are the eigenvectors of the covariance matrix of a set of training images [157] (the Karhunen-Loève transform), see Figure 15. This attempts to find a new representation with fewer numbers (dimensions) that still captures most of the variation in the sample without introducing any new confusions. A simple way of viewing these methods is that a compact representation of a face image is computed, which enables the formulation of distances between images that can be efficiently computed for $1:\ many$ matching.

Appearance methods can be global [5, 47, 88, 157] where the whole face is considered as a single entity, or local, where many representations of separate areas of the face are created. [144, 155, 175]. In many face recognition geometry and appearance are combined, and indeed to apply appearance-based methods in the presence of facial expression changes requires generating an expressionless "shape-free" face by image warping.



(a) (b) (c)

Figure 14: Approaches to face verification: a) geometric features; b) local appearance features; and, c) global appearance.

Figure 14 tries to clarify the various approaches to face recognition. Figure 14a shows the face image of one of the authors with the automatically detected geometric features superimposed. Eye features like pupil and eye corners are detected trough the glasses of the subject; features on the nose and the mouth are

Figure 15: A face image (left) can be decomposed as a sum of weighted Eigenfaces (images curtesy Turk & Pentland [157]).

also extracted. Figure 14b shows sample windows important face features and local face feature appearance models similar to the eigenface models for complete faces are often used. These are then "eigen-eyes", "eigen-noses", etc. Figure 14c shows the segmentation of the complete face from the image, which could be used for eigenface approaches to face recognition.

Considerable progress has been made in recent years, with much commercialization of face recognition, but a lot remains to be done towards the 'general' face recognition problem. To date, not much progress has been made towards modeling faces in 3D, while recognizing the face from images.

Despite all this work, face recognition is not sufficiently accurate yet to accomplish the large-population identification tasks tackled with fingerprint or iris. One clear limit is the similarity of appearance of identical twins, but determining the identity of two photographs of the same person is hindered by all of the following problems, which may be divided into four categories:

- *Physical face appearance changes:* These consist of quick changes in facial expression change, slow appearance changes because of aging and personal appearance changes. The latter may be due to the application of make-up, wearing of glasses, facial hair changes, changes in hairstyle and intentional disguises.

- *Acquisition geometry changes:* The appearance of the face in the image is in an unknown location has an unknown in-plane rotation and is of unknown size (scale). These are just the geometry changes when a person is looking straight in the camera. Rotations of a face in depth, i.e., facing the camera obliquely, introduce a host of differences of appearance of a face from one image to the next.

- *Changes in imaging conditions:* The lighting of a human's face can have large effects on the appearance of the face in an image, just consider for example front lighting against side lighting.

  Intrinsic camera characteristics and parameters may further change the appearance of a face in an image, quite independently of the ambient light. These camera characteristics include things like automatic white balancing and noise reduction.

- *Channel characteristics variations:* These are often quite unexpected image degradations because of compression-decompression artifacts. The commonly used compression standards like JPEG and MPEG are based on the compression of image blocks, and are not particularly designed shapes and forms of the human face appearance. This affects the performance of face recognition algorithms on compressed/archived data, such as legacy mug-shot databases and broadcast video.

No current face recognition system can claim to handle all of these problems well. For example, there has been little research on invariance of face recognition to aging effects. In general, to avoid the above

problems, constraints on the problem definition and image capture situation are used to limit the amount of invariance that needs to be afforded algorithmically. Indeed, it has been reported that automated face recognition systems can be used for matching mug-shot face images to face appearances in a crowd. These are positive biometric identification applications where there often may be a person in the loop who frames the face appearance in the image (i.e., does the "segmentation" part of the problem).

The main challenges of face recognition today are handling rotation in depth and broad lighting changes, together with personal appearance changes, i.e., to move beyond the mug shot scenario. Even under the best imaging conditions, however, recognition error rates need to be improved.

There is interest in other acquisition modalities such as 3D shape through stereo or range-finders; near infrared or facial thermograms [123], all of which have attractions, but lack the compelling advantages of visible-light face recognition that are outlined above. Particularly, the face image input devices are expensive.

4.3 Voice identification

Like face appearance, speaker identification [26, 56], sometimes referred to as "voiceprints" or *voice recognition* (& *talker recognition*) is attractive because of its prevalence in human communication and human day-to-day use. We expect to pick up the phone and be able to recognize someone by his or her voice after only a few words, although clearly the human brain is very good at exploiting context to narrow down the possibilities. That is, the topic of the conversation provides sometimes very significant cues about the identity of the speaker (though, we probably all have the experience of phone conversations where the context fails to contribute useful cues – "Hi, it's me.").

Figure 16: A piece of a voice signal, the signal has varying frequency content as a function of time t.

Telephony is the main target of speaker identification, since it is a domain with ubiquitous existing hardware where no other biometric can be used in a practical way at the moment. Increased security for applications such as telephone banking and "m-commerce" (commerce over the mobile phone) means the potential for deployment is very large. Speaking solely in order to be identified can be somewhat unnatural "active" authentication protocol), but in situations where the user is speaking anyway (e.g., a voice-controlled computer system, or when ordering something by phone) the biometric authentication protocol becomes "passive."

Physical and computer security by speaker authentication have received some attention, but here it is less natural and poorer performing than other biometrics. Speaker authentication is necessary for audio and video- indexing. Where a video signal is available, lip-motion identification has also been used [45, 89, 97].

Another description of speaker recognition can be found in [16] and we loosely quote some of this work here. It shows that opinions on a particular biometric can very from person to person.

> "Voice biometrics has a unique advantage over other biometrics because is relies on human speech, which is [like face] a primary modality in day-to-day, human communication, and provides a non-intrusive method for authentication. By extracting appropriate features from a person's voice and modeling the *voiceprint* [this is what the speaker community calls machine representation of a voice], the uniqueness of the physiology of the vocal tract and the articulatory voice properties can be used for recognizing the identity of a person. Recognizing a user based on voiceprints is commonly known as speaker recognition in the academic community, encompassing speaker verification, speaker identification, speaker classification, speaker segmentation and speaker clustering. Speaker recognition accuracy has improved significantly over the last few years, and a recent independent study compares speaker recognition favorably with respect to fingerprint recognition and other biometrics [102].
>
> When used in a text-independent mode (i.e., no constraints on the words to be spoken), voiceprint recognition offers many other advantages. Users do not have to remember passwords or pass phrases. Users do not have to go through a separate process for verification, since anything they say as part of the transaction dialog can be used to verify their identities, resulting in a truly integrated and non-intrusive verification process [authentication protocol]. Text independent verification is usually also language independent, and the user can speak in multiple languages or different languages for enrollment and authentication. Speaker recognition based on voiceprints, or acoustic speaker recognition, is appropriate for a wide variety of applications, especially remote authentication over legacy phone lines."

We can categorize speaker authentication systems depending on the freedom in what is spoken; this taxonomy based on increasingly complex tasks also corresponds to the sophistication of algorithms used and the progress in the art over time.

- *Fixed text:* The speaker says a predetermined word or phrase, which was recorded at enrollment. The word may be secret, so acts as a password, but once recorded a replay attack is easy, and re-enrollment is necessary to change the password.

- *Text dependent:* The speaker is prompted by the authentication system to say a specific thing. The machine aligns the utterance with the known text to determine the user. For this, enrollment is usually longer, but the prompted text can be changed at will. Limited systems (e.g., just using digit strings) are vulnerable to splicing-based replay attacks.

- *Text independent:* The speaker authentication system processes any utterance of the speaker. Here the speech can be task-oriented, so it is hard to acquire speech that also accomplishes the impostor's goal. Monitoring can be continuous, and the more that is said the greater the system's confidence in the identity of the user. Such systems can even authenticate a person when they switch language. The advent of trainable speech synthesis might enable attacks on this approach.

- *Conversational:* During authentication, the speech is recognized to verify identity by inquiring about "soft" knowledge that may not particular secret.

  False acceptance rates below 10^{-12} are claimed possible, making conversational biometrics very attractive for high security applications [98, 125].

Speaker authentication, however, suffers considerably from any variations in the microphone [61, 136] and transmission channel. Also, performance deteriorates badly when enrollment and use conditions are mismatched (for example, enrollment over a land line and authentication over a cell phone). This, of course, inevitably happens when a central server carries out speaker authentication from telephone signals. Background noise can also be a considerable problem in some circumstances, and variations in voice due to illness, emotion or aging are further problems that have received little study. Figure 16 shows a piece of a speech (voice) signal, in this case relatively clean distinct utterances of digits. It is not hard to imagine that this signal will be quite different if the speaker has (say) a cold.

Campbell in [26] enumerates a number of problems with voice recognition that we repeat in Table 6.

Misspoken or misread prompted phrases.
Extreme emotional states.
Time varying (intra- or intersession) microphone placement.
Poor or inconsistent room acoustics (e.g., multipath and noise).
Channel mismatch (e.g., using different microphones for enrollment and verification).
Sickness (e.g., head colds can alter the vocal tract).
Aging (the vocal tract can drift away from models with age).

Table 6: A voice biometric is very susceptible to the state of the subject and environmental issues.

Speaker verification is particularly vulnerable to replay attacks because of the ubiquity of sound recording and playback devices (see Section 13). Consequently more thought has been given, and should be given, in this domain to avoiding such attacks.

While traditionally used for verification, more recent technologies have started to address identification protocols, one particular domain being in audio and video indexing [4]. As noted above, recent and interesting development [98, 125], combining voiceprint recognition with the exchange of knowledge in an interactive authentication protocol (called *conversational biometrics*), can provide higher accuracy and is discussed further in Section 14.3 on dynamic protocols.

4.4 Iris identification

Although iris [172, 173] is a relatively new biometric, it has been shown to be very accurate and stable, mainly because there is no elastic distortion (except pupil dilation) from one sample to the next of the iris image as is found to be the case fingerprint images. The colored part of the eye bounded by the pupil and sclera is the iris, is extremely rich in texture and has proven to be a strong universal biometric identifier with great discriminating properties. So far, in the literature, there have been only a few iris recognition systems described. Perhaps the most well-known iris recognition system is the one designed by John Daugman [38, 39]. The iris texture is represented using Gabor wavelet responses and the matcher is an extremely simple and fast Hamming distance measure, as in Fig 18.

Like fingerprints, the appearance of this biometric identifier is the result of the developmental process in the womb and is not dictated by genetics. Quoting Daugman [39]: "Just as the striking visual similarity of identical twins reveals the genetic penetrance of overall facial appearance, a comparison of genetically identical irises reveals that iris is a phenotypic feature, not a genotypic feature."

For iris identification there has been very particular confusion about the *intrinsic error rate*, the theoretically possible lower bound, Section 8, and the practical, attainable error rates of an iris authentication

Figure 17: An iris image acquired under ideal circumstances (curtesy J. Daugman [38]).

application as described in Section 5. The experiments in [38] are performed on iris images obtained from an optometrist (see Figure 17), and show low FA and FR rates. The paper further derives lower bound on the error rates while making optimistic assumptions about the iris image quality and acquisition. It is unfortunate that many marketing claims have been based on the scientific paper [38].

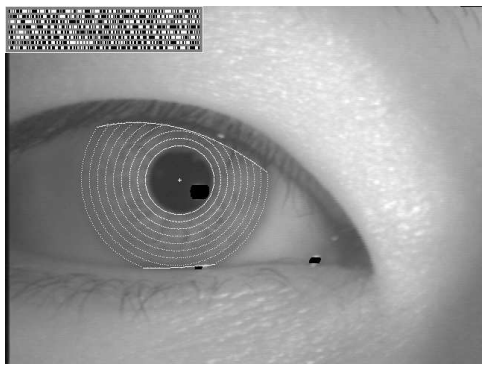


Figure 18: The iris codes are extracted from concentric circular strips around the pupil (curtesy R.P. Wildes [173]).

The primary problem in iris recognition is the design of an iris image capture device that is convenient and unobtrusive. It should be user friendly and yet capture the iris image with enough invariance from one ambient lighting situation to the next ambient lighting situation. A first step is to find the human face in the image using, for example, stereo techniques as described in [173]; then a high quality image of the iris needs to be acquired. The iris image capture device should further able to deal with specular reflection off the eyeballs; and with glasses and contact lenses (where the small hard contact lenses may create the most problems).

In summary, quite ingenious image acquisition devices are required to control the imaging of the iris and thus the FA, FR and FTE rates. In [15] an imaging setup is described to find objects close to the camera, this can be employed in the near infrared.

4.5 Hand geometry

There is no ideal biometric measurement; each biometrics has its strengths and limitations, and accordingly each particular biometric appeals to a particular authentication application. Suitability of a particular biometric to a specific application depends upon several factors [77, 75]; among these factors, the user acceptability is very significant. For many access control applications, like immigration, border control and dormitory meal plan access, very distinctive biometrics, e.g., fingerprint and iris, may not be acceptable for the sake of protecting an individual's privacy. In such situations, it could be desirable that the given biometric identifier be only distinctive enough for verification but not for identification.

Hand geometry, as the name suggests, refers to the geometric structure of the human hand, better yet, the geometric invariants of a hand. Typical features include length and width of the fingers, aspect ratio of the palm or fingers, width of the palm, thickness of the palm, etc. [13], see Figure 19. To our knowledge, the existing commercial systems do not take advantage of any non-geometric attributes of the hand, e.g., color of the skin.

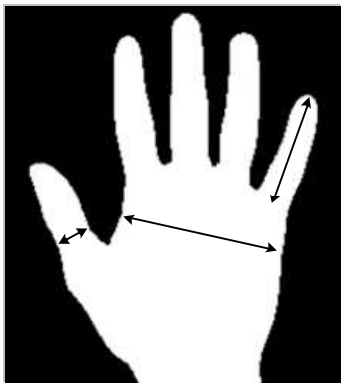


Figure 19: Examples of features that represent hand geometry.

Although these metrics do not vary significantly across the population, they can however be used to authenticate the identity of an individual. Hand geometry measurement is non-intrusive and it involves a simple processing of the resulting features. Unlike palm print verification methods [178], this method does not involve extraction of detailed features of the hand (for example, wrinkles on the skin).

Hand geometry-based verification systems are not new and have been available since the early 1970s. However, there is not much open literature addressing the research issues underlying hand geometry authentication; much of the literature is in the form of patents [49, 69] or application-oriented description [106]. A notable exception is the prototype system described by Jain et al. [74]. Sidlauskas [147] discusses a 3D hand profile identification apparatus that has been used for hand geometry recognition.

Hand geometry authentication is also attractive for various other reasons. Almost all of the working population has hands and exception processing for people with disabilities could be easily engineered [179]. Hand geometry measurements are easily collectible, compared to, say, iris and retina, due to both the dexterity of the hand and due to a relatively simple method of sensing, which does not impose undue requirements on the imaging optics. (As described in Section 7 and [74] these systems however may have some usability problems that need further research.) Note that good frictional skin is required by fingerprint imaging systems, and a special illumination setup is needed for the above-mentioned biometrics such as iris and retina.

Hand geometry does not have these restrictions. As the computations are also fairly simple, a stand-alone system is easy to build. Further, hand geometry is ideally suited for integration with other biometrics, in particular, fingerprints. For instance, an authentication system may use fingerprints for more precise authentication and use hand geometry for less stringent authentication. It is easy to conceptualize a sensing system, which can simultaneously capture both fingerprints and hand geometry.

In sum, authentication of identity of an individual based on a set of hand features is expected to play a role in biometric authentication. It is well known that the individual hand features themselves are not very descriptive and that hand geometry authentication has relatively high FA and FR. Devising methods to combine these non-salient individual features to attain robust positive identification is a challenging pattern recognition problem in its own right.

4.6 Signature verification

Signature verification (e.g., [112]) is another biometric that has a long pedigree before the advent of computers, wide usage in document authentication and transaction authorization in the form of checks and credit card receipts. Signature recognition is an instance of writer recognition, which has been accepted as irrefutable evidence in courts of law. Signatures also come in a wide variety, see Figure 20, thereby giving the signatory the ability to choose the "distinctiveness" and "uniqueness" of the signature, which will influence his or her FA and FR rate (see Section 12 on sheep and wolves).

Figure 20: Signatures come in a wide variety.

The natural division of automated signature verification is by distinguishing the technologies by the sensing modality, i.e., on-line vs. off-line:.

1. *Off-line* or "static" signatures are scanned from paper documents where they were written in the conventional way [121]. The lack of further information about the signature acquisition makes these techniques very vulnerable to forgery.

   Incidentally, the problem of *writer authentication* (e.g., [65]) always presents itself in the form of analyzing scanned paper documents and falls in this class. Approaches to writer authentication (and signature verification) are typically based on features, such as, *number of interior contours* and *number of vertical slope components* [151].

2. *On-line* or "dynamic" signatures are written with an electronically instrumented device and the dynamic information (pen tip location through time) is usually available at high resolution, even when the pen is not in contact with the paper.

   Approaches to dynamic signature verification include signature representations that are based on Euclidean distances and regional correlation matching measures, or probabilistic representations such as Hidden Markov Models (see, e.g., ([42]).

Consequently, the technical approaches can be divided into approaches based on temporal features and approaches based on spatial features; few formal verification procedures have developed in signature verification technology, which we will become clear in Section 7.

A first comprehensive approach to static signature verification can be traced back to [111]. The development of dynamic signature capture devices (see, e.g., [46] resulted in much activity because a notion of time beyond two-dimensional space (paper) was introduced, for instance $x(t)$ and $y(t)$, the location of the pen while signing. The devices record a stream of 5-dimensional vectors $(x, y, p, \theta_x, \theta_y)$ sampled equidistant in time; here p is the axial pen force; and the θ_x and θ_y describe the angles of the pen with the X-Y plane [42]).

The signature further has an interesting way of looking upon false accepts in terms of defining the sophistication level of the forger (attacker), in categories like *zero-effort forgery, home-improved forgery, over-the-shoulder forger,* and *professional forgery*. There is no clear definition of level of sophistication of impersonalizing for any other biometric.

Because of this special hardware needed for the more robust on-line recognition, it seems unlikely that signature verification will spread beyond the domains where it is already used, but the volume of signature authorized transactions today is huge, making automation through signature verification very important. Moreover, signatures are often captured electronically already, merely to reduce paper storage and transport. There are approaches for pen location and orientation estimation using visual light being researched [110]. This may eventually lower the cost of signature acquisition and may, even, influence ideas toward three-dimensional signatures.

As determined in [31], and already mentioned, characteristics of a biometrics are (i) universality, (ii) uniqueness, (iii) *permanence*, meaning invariance over time, and (iv) collectability. The characteristic of permanence of signature is questionable since a person can change his or her signature pretty much at will at any point in time. (A person could even insist on a different signature for every day of the week.) In a sense, the genetically and environmentally determined muscle dexterity of the hand is translated into a visual and machine-readable token. Hence, this biometric is (as face and voice) affected by illness, emotion or aging that have received little study.

5 Quantitative parameters

There are a number of biometric application parameters that need to be specified before a particular biometric is selected and even before a solution is designed. The first ones to look at are the FA and FR rates for a biometric authentication application.

These parameters are statistical properties of the authentication application and are optimized and tested as described in the sections on training and testing (Sections 11.1 and 11.2). In this section, we present more intuitive and theoretical meanings associated with these parameters.

A first thing, however, that is confusing and we mentioned before but is worthwhile to point out here, is shown in Figure 21. This is the use of the terms *False Positive* FP and *False Negative* FN. If one wishes, it is perfectly OK to use these terms interchangeably with *False Accept* FA and *False Reject* FR.

However, in this document, we will use the terms FP and FN for "screening" (identification) applications *only*: A false positive FP is the erroneous decision that an identity is in the database, while in reality it is not. A false negative FN is the erroneous decision that an identity is *not* in the database, while in reality it is in the database.

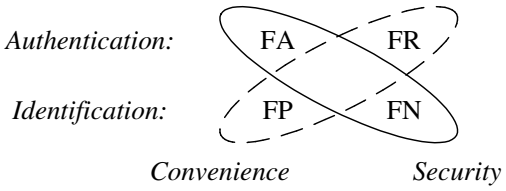


Figure 21: The terminology is confusing.

On the other hand, we use the terms *False Accept* FA and *False Reject* FR exclusively for authentication applications. As shown in Figure 21, false accepts and false negatives are security issues while false rejects and false positives are convenience issues. (Technically however a FA is the same as a FP and a FR is the same as a FN.)

5.1 Error rates

Determining the probability of a match of two biometric samples is an important biometric task and the precision with which this can be achieved is influenced by several factors. The variability in the biometrics input signal is unfortunately much higher from one sampling to the next as compared to the variability in (say) passwords due to input (typing) error.

The probability of a match is the probability that the submitted biometric sample matches (in some sense) a stored reference sample. Often this probability is translated into a score s. In a biometrics system, one can explicitly set a threshold t on this score to directly control the FA and FR rates by authenticating a match if $s \geq t$. Given a high matching score, the system can guarantee that the probability of the signal coming from a genuine person is significantly high and the probability that the signal is coming from an imposter is significantly low (low FA).

Conversely, given a low match score, the probability that the sample is genuine is low (low FR) and the probability that the sample is an imposter is high. Setting the match decision threshold t low guarantees low FR at the expense of higher FA.

This is shown in Figure 22, the *imposter scores* tend to be low while the *genuine scores* tend to be high. The threshold t controls the tradeoff between FA and FR.

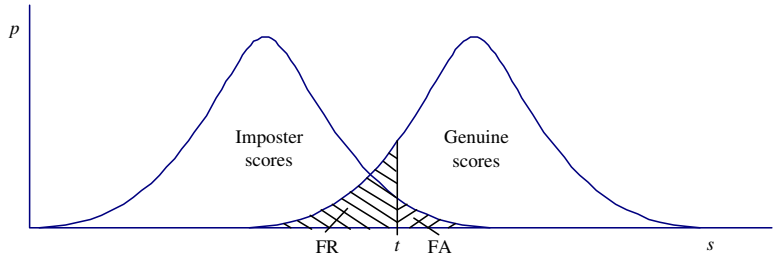


Figure 22: The imposter scores are on average lower than the genuine scores.

5.2 Probability of match

As noted above, the authentication decisions related to a match score s should be based on *both*, the probability that the biometrics input sample matches the reference sample *and* the probability that the input sample *does not* match the reference sample. Given the score densities as in Figure 22, ideally for a given score s one would like to compute the probability of a FA and the probability of a FR. However, the probability of FA and the probability of FR, given a match score s cannot be directly measured but only be estimated. The estimation of probabilities of FA and FR entails training the matcher on test databases of biometric samples (see Section 11.1). Consequently, probabilities of FA and FR for a given decision (about score s) can only be estimated if either the training data represents the target population (e.g., samples associated with score s) well and there exists enough training data or if the expected subject population can be tractably modeled from the training data.

The probability of match, which is roughly $2 - \mathrm{FR}$ if $FA \ll FR$, is related to the match score s. Much more is known about the population, or customers, of an application than is known about the enemies. Consequently, the probability of a FA, a false match, is hard to estimate. Hence, the FR rate for a particular decision is easier to estimate than the FA rate for that decision, because the biometric samples of the enemy population are not available. (An interesting solution to this is the use of *"cohorts"*, see Section 11.1.1).

Along with the match (and mismatch) probabilities, ideally measures of confidence (Section 11.2) should also be given. This allows the application to make an informed decision (using decision theory, risk analysis, etc.). Typical training data consists of significantly fewer samples for match scores than those for non-match scores. Consequently, the confidence in the FR is typically much lower than that in an FA.

This probability of match can be used to counter repudiation efforts, which is a difficult issue when passwords are used. Here repudiation is the denial of a transaction by a subject. The probability of match is also used by the application to arrive at an authentication decision. This probability is computed by a matcher and the probability is transmitted to the application.

5.3 Tradeoffs and operating points

Most biometric authentication systems cannot guarantee error rates that are low enough for applications that have to be both secure (low FA) and convenient (low FR), see Section 6. Given the match score, what can be guaranteed is that for any FA there is a corresponding upper bound on the FR and *vice versa*. That is,

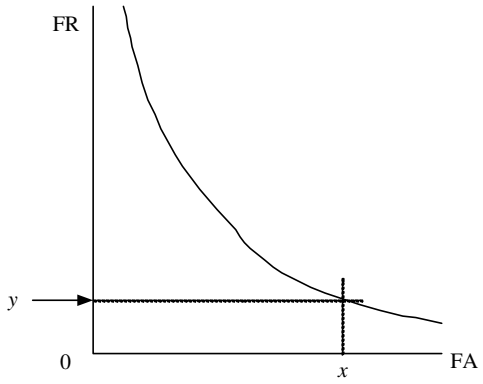


Figure 23: The ROC expresses the tradeoff between FA and FR.

the FA and FR behavior is expressed in terms of a Receiver Operator Characteristics (ROC) curve [57]. An example of an ROC curve is shown in Figure 23, for example, operating at $\mathrm{FA} = x$ means that the $\mathrm{FR} \geq y$.

Note that by excluding some of the most difficult data (e.g., the "goats," the hard to match subjects, see Section 12), the overall ROC can be improved. Usually this is an acceptable practice as long as the percentage excluded is not too great. For instance, asking for a "retry" on a fingerprint image 1% of the time is not unreasonable. However, excluding 1% of the *people* may not be tolerable.

The question to ask is what happens if the excluded subjects (the 1% of the subjects) are allowed to enroll in the biometric authentication system. Let us look at this in the context of fingerprint and minutiae matching. In some cases an excluded subject would contribute to the FR rate (e.g., a fingerprint with no extracted minutiae could never be matched). In other cases it would contribute to the FA rate (e.g., almost any fingerprint would match to a template with a single minutiae). In general, if the tails of the ROC curve do not asymptote at zero FA and zero FR, there is probably some data that could be profitably excluded. The trick is finding some automatic way of detecting these poor data items (see the section on sample quality control, Section 12.4).

Since tradeoff between the FA and FR rates largely expresses itself as a security versus convenience tradeoff (see 6), when designing a biometric authentication system, the first question that should perhaps be asked is: "For this application, is security of prime concern, or is convenience the real issue in this application?" The latter would, for example, be the case in voluntary applications because there the convenience may be the *deciding* factor in the success of a particular installation. In that case, as shown in Figure 23, convenience could be selected at some $\mathrm{FR} = y$ and the corresponding (lack of) security is expressed as $\mathrm{FA} \geq x$. When security is most important, security can be fixed as $\mathrm{FA} = x$, which implies some FR.

This means that using the same biometrics matcher for both a secure application and a related convenient application may not necessarily be the optimal solution. This is, for example, the case for the ROC curves in Figure 24a. Here the ROC curve for Matcher A corresponds to a more secure matcher since, for low FA, the achievable minimum FR is lower than that for ROC curve of Matcher B. In that operating area, where the inconvenience expressed as FR is relatively high (higher false reject rate), Matcher A has better characteristics than Matcher B. Matcher B may be selected instead if low FR, convenience, is preferred over security.

Typically, a system is designed by selecting a FA (for security) or FR (for convenience) and hence only

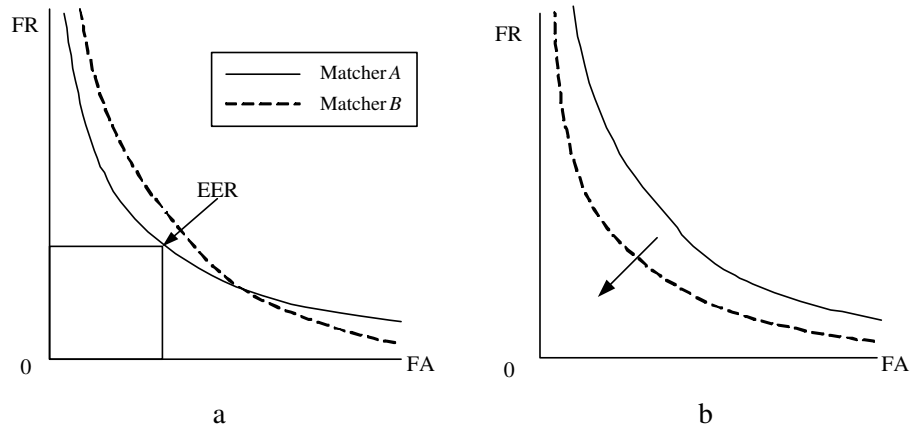


Figure 24: ROCs for Matcher A and B: (a) One matcher may be preferred for convenience and another for security; (b) The ROC can often be improved by excluding some data.

a portion of ROC curve is important. A particular biometric system is designed at one point on the curve, another system may be designed at another point on the curve. Therefore, a single number that somehow expresses the accuracy of a matcher such as the equal error rate (EER in Figure 24a) or d' [37] may not be that relevant. The EER is defined as the error at *that* operating point s where

$$\text{EER} = \text{FA}(s) = \text{FR}(s).$$

However, the operating point does not necessarily need to be static. For example, screening processes such as at airports can be multi-modal, in the sense that demographics, text, sound, image and video can all be used. Depending on demographic match, biometrics authentication can operate at different operating points on the ROC curve. For example, it may be the case that wanted person X is from state Y. One can then set the FA and FR rates differently when matching with a subject Z from state Y. That is, for all subjects from state Y the matcher will be operated at a lower FR when matching with wanted per X. This is an example of a dynamic authentication protocol.

5.4 Other error rates

Next to the issue of security *versus* convenience, there are many factors associated with biometrics applications. These include voluntary versus involuntary applications, habituated versus non-habituated users, and supervised versus unsupervised acquisition of the biometrics sample [71]. Section 10 discusses qualitative properties of applications and biometrics like that.

There are a number of quantitative variables in addition to the well known FA and FR rates. Some of these are listed in Table 7.

There is the *failure to acquire* (FTA) rate which is the percentage of the target population that does not posses a particular biometrics, i.e., which do not deliver a usable biometric sample. Another variable is the *failure to enroll* (FTE) rate, which is the proportion of the population that somehow cannot be enrolled because of limitations of the technology or procedural problems [11]. Here what the exact definition of a FTA is subtle. It can be that a subject does not posses the biometrics that is needed for enrollment, i.e., the subject is missing an eye; or, it can be that a subject's biometrics cannot be measured, say, the fingerprint

FA	False accept of an intruder causing security problems.
FR	False reject of an authorized causing inconvenience.
FTA	Failure to acquire causing a failure to enroll.
FTE	Failure to enroll causing failure to use.
FTU	Failure to use – is a significant cost factor any biometrics application.

Table 7: Quantitative variables of a biometrics application.

of a brick layer (the ridges have been worn away). Technology may well be improved so that this particular subject can be enrolled at some future point. Both FTA and FTE are partially due to intrinsic biometrics properties and limitations in the biometrics state of the art.

We introduce another application variable, mainly for voluntary applications, the "failure to use" FTU rate (also a random variable, at least during the design stage of the application). This is FTE rate plus the proportion of the population that for some reason does not enroll, or enrolls and fails to continue using the biometrics system.

For voluntary biometrics authentication applications, the difference between the FTU and FTE rates will be due to convenience problems with the voluntary applications. For involuntary applications, these rates are in theory the same (if it is affordable). There exists some intrinsic lower bound on the FTA for each biometrics because some portion of the population, cannot show, or does not posses the particular biometrics. This is sometimes referred to as a biometrics' *universality* [71] issue. The FTE rate, on the other hand, can be used as a system parameter to design and build an installation within budget. The FTE can also be controlled by biometric sample quality control, Section 12.4.

For a given application, a particular FA, FR operating point needs to be determined. This operating point, obviously, is hard to establish beforehand. Even when an installation is in place, the genuine appearance of an FA may never be detected. For a voluntary application, the FTU rate, of course, can be determined when the installation is in place. The technical reasons for a high FTU rate will most probably be convenience (usability) problems (although there may also be non-technical reasons). For an involuntary application, on the other hand, the FTE rate can be measured, which is usually done in laboratory-type (e.g., more supervised, controlled) situations. Tests of biometrics authentication generally use subjects on a voluntary basis and it is in general unclear how the FTU (which could be interpreted as not volunteering) influences the FTE. On the other hand, given a database of test samples, by increasing the FTE the corresponding ROC curve can be made better and better. That is, as mentioned before, by removing undesirable samples, ROC curve **A** in Figure 24b can be transformed into almost any ROC curve **B** that is closer to the coordinate axes.

What the ROC of an operational biometrics installation is, and at what point the system is operating, will depend entirely on the enrolled population and the desired security. The design parameter FTE can be artificially increased for a given installation, which will improve the overall quality of the enrolled population at the expense of increased exception handling. The system variable FTE enables a tradeoff between *manual versus automated* authentication, which in turn is related to cost.

6 Qualitative properties

This section touches on three properties of automated biometric authentication systems and hence three properties that can be associated with the various biometrics. First we look at the security and convenience of an implementation (biometric). Security is often associated with low FA rates, while convenience is associated with low FR rates. Hence, the tradeoff between FA and FR translates into a tradeoff between security and convenience. By setting FA $= 1$, we have FR $= 0$, a very convenient but insecure implementation, this relates to an authentication protocol for accessing churches and other such public spaces. Setting FA $= 0$ and FR $= 1$, on the other hand we have a closed system.

A third property of biometric authentication is privacy of the application (biometric), which is of course very much related to its security. What privacy means from an information technology point of view is poorly understood end here we try to open the beginning of a discussion.

6.1 Security

As indicated in Section 3, biometrics is related to authentication protocols and policies and enrolment policies, which are traditional security topics. It therefore seems self-evident that the area of computer security should embrace biometrics. From a security point of view, non-repudiation [152] is biometrics' main strength, i.e., because the user transmits biometrics information to the application that is unique to the user, the user cannot deny that the application was used or accessed. This, in itself, adds deterrence over traditional authentication. Non-repudiation, incidentally, is a well-defined computer security property of authentication systems [152]. When biometrics are concerned, non-repudiation is guaranteed only with certain statistical probabilities though.

Biometrics distinguishes itself from traditional authentication protocols in various ways and perhaps biometrics' most notorious liability is *impersonalization* (see Section 13.2), which is also perceived as a gross security flaw of biometric technology. Impersonalization may appear quite simple like lifting fingerprints of glasses, to quite brutal like cutting off fingers. Most everyone is familiar with these types of things. This fiction about biometrics, along with the legacy law enforcement uses, affects the acceptance (Section 10) from one biometric identifier to the next. These acceptance problems are in essence due to the fact that impersonating a biometrics has a stigma associated with it because it amounts to violating the identity associated with the biometrics. It should be noted, however, that lifting a latent fingerprint and using the print to impersonate others is in essence the same as covertly observing password input by say hidden cameras. However, there is a large difference in the difficulty of applying knowledge about a biometric as opposed to applying knowledge about a password to gain unauthorized access. This is still the case if a fingerprint scanner can be easily fooled by fake biometrics, it is never as easy as impersonating an identity by using a stolen password.

Biometric authentication does not guarantee, and cannot guarantee, 100% certainty of the matching decision. Automated authentication using other credentials or tokens is done by exact matching but, in turn, there is no guarantee that the token is *not* in the possession of impersonator. Perhaps the most pressing issues with security of biometrics are:

- Tight integration of the traditional authentication protocol and the biometric authentication. All too often, biometrics are introduced into an application as an afterthought.
- Automated biometric authentication will be used for authentication at large scale. The field of security research, sooner or later, will have to adopt biometrics.

- Biometrics brings the need to analyze security in probabilistic terms. Probabilistic frameworks for integrated enrollment with traditional authentication and biometric authentication need to be developed.
- Since automated biometric authentication does not provide a *'YES/NO'* answer, but rather an answer in terms of probabilities, things like risk analysis can be used during authentication. This analysis should be part of the authentication protocol (Section 3).

6.1.1 What is security

Security is the protection of an application against threats. There is the threat of a nuclear missile attack on the system. There is the threat of certain people impersonating other people (impersonate the users of the system or other individuals) to somehow violate or attack the application or the users of the application. Securing an application against people is implemented with an authentication system that is more or less automated with various authentication methods (combinations of possessions, knowledge or biometrics).

Hence, we have here unfortunately that the very object (people) of the application are also the threat to the application and are the main target of the application too. This is what makes the transportation application so different from a financial application, where the object and target of the application is money. The biometrics is intended to protect the application against people, the enemies of the application in particular.

Here biometric authentication systems are inherently more secure than legacy authentication systems because there is, in theory, a "more secure" linking of subjects to the universally better accepted identity databases, universal databases (passports, birth certificates).

In any biometrics authentication system, the weakest point is the most serious vulnerability since this is the easiest attack point. This is restated as: "Principle of Easiest Access: An intruder can be expected to use any available means to access the application and it cannot be expected that the most protected points of access will be attacked." The key here is that introducing biometrics does not create novel vulnerabilities and security loopholes. This means that for any biometrics authentication system where security is important, the introduction of a biometrics has to make sense and biometrics should be an integrated aspect of the overall security of the application. Note that the security of biometrics systems is related to protection against career criminals; if career criminals are not an issue, the biometrics can be used to improve convenience of an application. See the below section on convenience.

6.1.2 Threats

People are threats but of course the vast majority of people just are interested in the application in itself, access to a locale, a privilege, or a service. The enemies of the application and of the authentication process are divided into three groups [118]:

- *Amateurs:* These are subjects, probably largely users of the application, who observe a security flaw in the system and start making use of this flaw.
- *Hackers:* Subjects that attempt to access the system, or attack the authentication system itself. The intent is not to hurt anyone or steal anything, it is just to see if it can be done.
- *Career criminals:* Subjects that attack the system for malicious purposes, i.e., the threat of criminals using the privileges and/or gaining access to the system. Such career criminals understand the targets and vulnerabilities of the application.

This is a categorization as can be found in text books on information security (i.e., [152]). Translating this to the domain of physical security, and taking an airport application, the last category, the career criminals, are the real threat, i.e.,

- *Terrorists:* Subjects that attack the system for malicious purposes. This corresponds to the threat of individuals using and abusing the application for a terrorist attack of the application or a terrorist attack using the application.

Clearly, the terrorist threat means that the main biometric threats are *circumvention* and *impersonation* (Section 13.2) as outlined in the section on *biometric attacks*.

Translating the population groups of *amateurs* and *hackers* to the domain of physical security, the *only correct way* to handle these groups is to merge them with the enemies of the application, the career criminals, the hijacker, the terrorist, and so on. There seems no way around it, any way of defining a third group will have undesirable security implications because there is no way to construct such a group. Therefore, this leaves us with two population groups, *friends* and *foes*.

Consequently, the only secure implementation of biometric authentication system is to enroll all subjects that form no threat to the application by mandatory enrollment including screening, and not to allow any other subject to the application. This is obviously impossible and undesirable and therefore subjects need to be enrolled in probabilistic way anyway, even if no biometric is used in the authentication protocol. Section 12 discusses this issue when dealing with authentication database enrollment. Hence, the main point of this section is that security is already associated with FA rates but more fundamentally to the quality of enrollment, which can be expressed in terms of probabilities.

6.2 Convenience

When talking about "convenience," the biometrics literature is sometimes somewhat confusing, in that really two ideas are described by convenience:

- *The convenience of a biometric:* This is a somewhat nebulous concept of an intrinsic user-friendliness of a biometric. This is expressed by properties of biometric identifiers as described in Section 10.

  The problem is that the *natural biometrics* and most convenient biometrics like face and voice are also weak biometrics (Section 8). Convenience hence involves a tradeoff between the properties of the biometrics, in that, with the possible exception of the iris biometric, the more accurate biometrics are also less convenient and natural.

- *The convenience of a particular implementation:* This is the ease with which a correctly authorized person is authenticated on access of the application. This includes things like availability, the workflow of the authentication process, the exception handling process, and the false rejects (FR) of the authentication (and the false positives (FP) of the screening, Section 3.8).

Notwithstanding, the FR rate is often used as some measure of convenience of an authentication application, e.g.,

$$\text{"Convenience"} = (1 - \text{FR}),$$

the higher the FR, the less convenient an application is because more subjects are incorrectly rejected and therefore subject to the exception handling process.

Similarly, the FA is often used as a measure of security of a biometrics authentication application, i.e., the lower the FA, the better the security. Loosely this is expressed as

$$\text{"Security"} = (1 - \text{FA}).$$

Hence, security is a tradeoff with convenience in any biometric authentication system as shown in Figure 25: e.g., setting FR $= 1$ gives FA $= 0$ and hence perfect security but closed system. But when as in the figure the desired security level (or FA) is set, the associated convenience, expressed as $(1 - \text{FR})$ is pretty much a given.

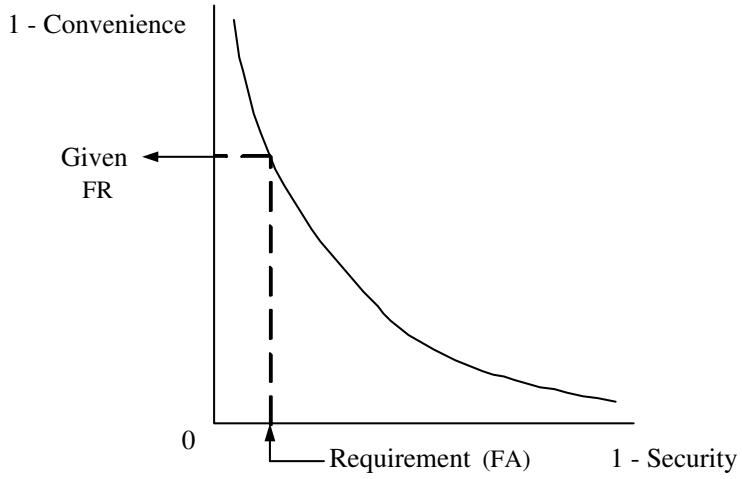


Figure 25: To some extend, the ROC expresses the security *vs.* convenience tradeoff.

What exactly the value of $(1 - \text{FR})$ or loosely the convenience of an installation is, of course, another question. It is hard to establish what the associated convenience level (in terms of FR) is in some formula. The only way to design a biometric authentication system therefore is to design realistic statistical tests. Testing is discussed in Section 11.2.

For screening systems $\text{"Security"} = (1 - \text{FN})$, i.e., the lower the chances of missing undesirable subjects in the databases, the higher the security. The convenience would be given by $\text{"Convenience"} = (1 - \text{FP})$, where the inconvenience is to the falsely matched people who will be subject to further interrogation. When the FR and FN rates become too high, the application becomes inconvenient to all users, since the access control point will no doubt will be congested. In Section 8, we quote some dramatic numbers from [114] related to high FP rates of biometric identifiers in a pure identification application.

An important thing to remember is that, as pointed out in Section 5, often the FR of biometric authentication systems is given without taking the FTE (Failure to Enroll) and sometimes not even the FTA (Failure to Acquire) into account. Using $(1 - \text{FR})$ as a measure of "convenience" then is overly optimistic and misleading.

6.3 Privacy

If not carefully implemented, biometric authentication may be fooled in many ways by the presentation, or the acquisition/ transmission of biometric identifiers for the enrollment of new identities (Section 12) or accessing applications through the generation of a biometric identifier (Section 13). Biometric authentication

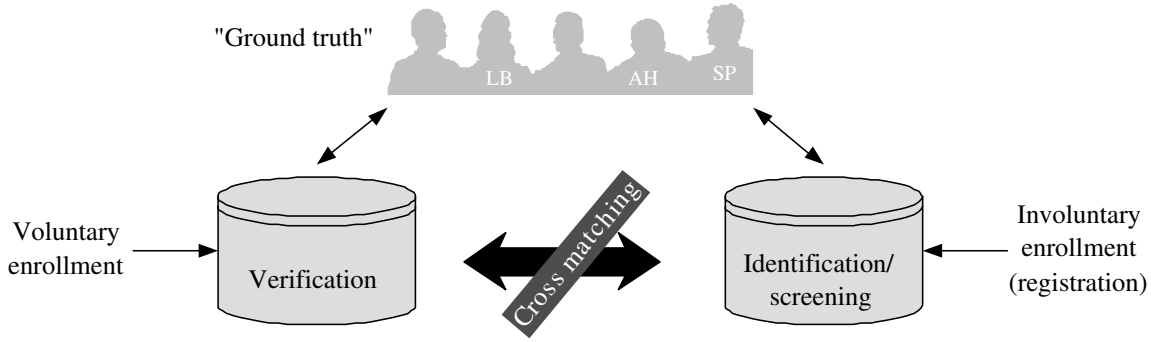


Figure 26: Large databases that can be cross matched are a privacy concern.

further comes with the dilemma that once a biometric identifier is somehow compromised, the identifier is compromised forever, which leads to security and privacy concerns.

Automated biometrics authentication are being developed, field tested, and installed for various larger scale financial access and physical access control applications. Consequently, privacy issues (data confidentiality issues, in security terms) with biometric authentication have been a topic of concern for quite some time now, also in the security literature [17, 128].

Here we really have two related issues [71]:

Privacy: Any biometric technology is traditionally perceived as dehumanizing and as a threat to privacy rights. As biometric technology becomes more and more foolproof, the process of getting authenticated itself leaves trails of undeniable private information, e.g., where is an individual? what is the individual buying? etc.

In case of biometric authentication, this problem is even more serious because the biometric features may additionally inform others about the medical history or susceptibilities of a subject, e.g., retinal vasculature may divulge information about diabetes or hypertension [104].

Consequently, there is a legitimate concern about privacy issues associated with biometric authentication.

Proscription: When a biometric measurement is offered to a given system, the information contained in it should not be used for any other purpose than its intended use. In any (networked) information processing system, it is difficult to ensure that the biometric measurement will only be used for intended purposes. Which may be hard to enforce, see Figure 26 where it is shown how easily civilian databases of trusted individuals can be linked to criminal databases.

Privacy and proscription concerns are summarized as follows:

1. Much data about people is already collected. There is concern about every bit of additional information that is stored about people, especially when it involves personal traits like biometrics.

2. Traditional security issues like *data integrity* and *data confidentiality* that may lead to violations of personal information.

3. Biometrics databases can be used for cross matching of databases. For example, matching against law enforcement databases, such as, the FBI or INS databases. This is a proscription issue and becomes a real problem when authentication databases are matched against legacy criminal databases, Figure 26.

These concerns are aggravated by the fact that a biometrics cannot be changed. One of the properties that make biometrics so attractive for authentication purposes, the invariance over time (permanence), is also one of the liabilities. When a credit card number is somehow compromised, the issuing bank can just assign the customer a new credit card number. When a biometrics is compromised, however, a new one cannot be issued. A techniques called cancellable biometrics that may alleviate privacy concerns is described in [131].

6.3.1 Questions

In general, many unanswered questions about *how* to make biometric authentication work without creating additional security loopholes, and without infringing on civil liberties, need to be answered. For national authentication applications, like travel, difficult questions will have to be answered about who will be eligible and therefore who will be enrolled in automated authentication systems. Other questions that need to be answered are who will administer and maintain databases of authorized subjects and how the data integrity of these databases is protected. Perhaps the largest issue with data integrity is keeping the databases ”clean” with strict criteria for enrollment and strict criteria for continued enrollment of subjects.

A new government bureaucracy may be created to maintain the enrollment database but cross matching as in Figure 26 is then a real possibility, especially for keeping the authentication database current. This matching of database based on voluntary enrollment with databases collected through involuntary means, such as criminal databases, may be very controversial. Therefore, another possibility is that one or more private entities maintain one or more of these databases and keep them current, very much like today’s system of credit rating. The question then is how these databases are kept current, which can only be achieved by somehow monitoring the behavior of the enrolled subjects.

When it comes to security, especially physical security, most everyone seems to be willing to give up some privacy. Privacy concerns are mainly with *information security*, i.e., data confidentiality and integrity and have to do with trust in the application, the technology and the authentication system operator. Unauthorized access of biometric authentication databases may lead to ”privacy attacks” with very undesirable consequences. The use of authentication methods will lead to impersonation attacks (Section 13.2) that can be viewed as the ultimate invasion of privacy. That is, the stealing of identities, which is a more personal violation when biometric samples are also stolen.

7 Realistic error estimates

There certainly is no shortage of performance and accuracy numbers out there. Biometrics is an emerging technology with fierce competition and manufacturers of biometrics systems are continuously refining the technology and claiming high accuracy. This may be defined simply as a system that works, or as a system that makes no or very few errors, or a system that is 100% accuracy. Obviously, such loose definitions of accuracy are undesirable and there is a need for a precise definition; "accuracy" is defined Section 5.

7.1 Testing scenarios

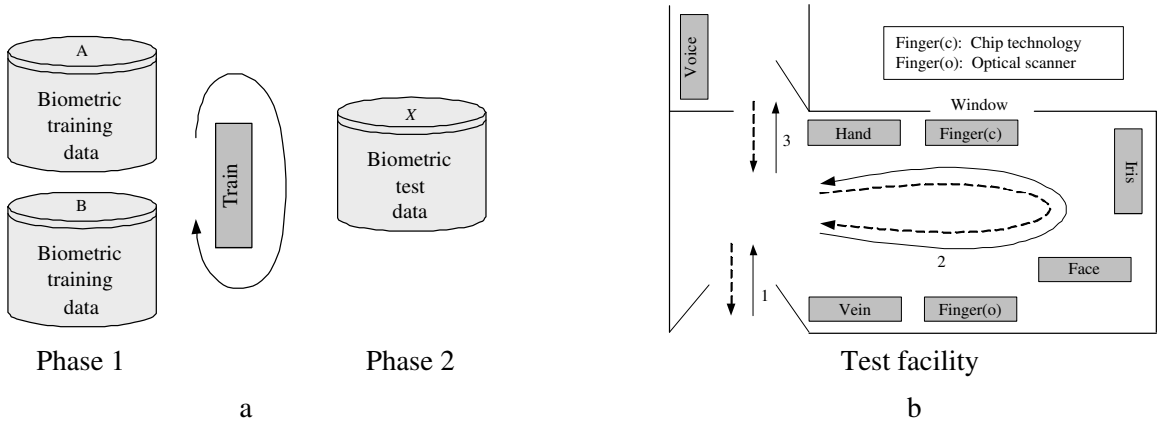


Figure 27: Two test scenarios for biometric matchers and systems: a) technology evaluation; b) scenario evaluation.

In this section, we look at accuracy numbers for our biometrics as much as we can reconstruct them from from the biometric literature. However, we should first distinguish between two testing methodologies for vendor evaluation are shown in Figure 27.

a: This methodology is depicted in Figure 27a: Databases A, B, ... with correctly labeled biometric samples are made available at the start of Phase 1 of the test. The samples are labeled in such a way that it is known, at least, what set of samples belongs to which subject. Hence, it is known which samples should match and which one should not match. (If the data are incorrectly labeled, the database is sometimes called *not clean*.)

The contestants in such a competition (often a mixture between commercial vendors and academia) have a certain period of time to train their algorithms on the data A, B, ... and submit their algorithms to an independent testing organization. This organization compares the different algorithms on a newly made available database X.

b: A scenario evaluation testing facility, e.g., is shown in Figure 27b. The biometric authentication devices are installed in an "office environment" as in the figure (loosely the test facility layout used in [103]) and the devices are tested by a group of volunteers over a period of time.

These voluntary test subjects authenticate themselves on the multiple biometric authentication systems in the test facility like the one in Figure 27b on a regular basis.

The scenario evaluations using end-to-end biometric authentication systems [102] are very different from evaluations found in [100, 124, 14], technology evaluations (type **a**) where biometric matchers are tested on databases (possibly collected from volunteers) though they have in common that biometric samples are mostly collected from volunteers. Figure 27 shows the difference in procedures; it is unclear how the test results obtained with such disparate testing procedures are to be compared.

The main problem with scenarios of type **b** is that such tests is not "*double blind*", which may greatly influence the results. For a double-blind test it is necessary that *both* the subject *and* the system (operators) do not know the answer returned. The test scenario **b** is of course everything but a double-blind test [33]; this simple fact would, in case of medical studies, invalidate the whole test. A true double-blind biometric authentication test has requirements that can be easily envisioned:

1. The biometric authentication application should be mandatory;
2. The subjects themselves should be unaware of the testing;
3. The user population should be fairly represented in the test;

and one can surely think of more. The implication is that, at the minimum, that these type of test should be performed under the most realistic circumstances and maybe the only place to perform biometric testing is *at the actual installation* (*operational evaluation*).

The biometrics face, finger, and voice are ahead in terms of technology evaluations (type **a**). Public are organized by universities (e.g., the University of Bologna) or by government agencies such as NIST and DoD [124, 14] and comparative evaluation reports are available for finger [100], voice, and face.

On the other hand, for estimating error rates of commercial and university authentication systems using iris and hand authentication, no public databases are available and no authentication competitions are being held held. Therefore we have to resort to public reports on biometric technology commissioned by government agencies or other institutions, these are mainly scenario evaluations (type **b**).

A separate section, Section 7.5, is devoted to signature verification error rates. Much of the testing is done with technology evaluations **a** but not is a very systematic way. There are also static off-line and dynamic on-line technologies but the results are scattered throughout the literature. Static approaches are published in *document analysis* while dynamic approaches are published in *signal processing* or time signal analysis literature. Dynamic signature is a biometric that lends itself well to test scenario **b** but not no such tests are reported in the literature.

7.2 Implications of error rates

Before we look at actual numbers for the various biometrics, let us look at the implications of various values for these error rates. O'Gorman [115] gives two examples below that point out problems with the state of the art.

"Biometric Authentication – Why does it reject me?"

Verification protocol: Frequent flyer smartcard with biometric:

1. Assume a system where each person is verified with a fingerprint template on a smartcard or stored in a central database.

2. Hence, each passenger uses a unique frequent flyer number and a fingerprint sample as credentials.

3. Use a best-case false reject (FR) rate for finger of: 0.03 (= 3%).

4. If 5000 people per hour are requesting access (Newark airport hourly volume) in a 14 hour day, roughly *2100 people will fail to be verified (FR)*
   (3% of $5000 \times 14 = .03 \times 70000 = 2100$).
   Of course one could pick a different operating point (e.g. FR = 0.3%).

5. These people have to be verified through some exception handling procedure (which is part of the protocol). Note that this might be as simple as a retry in some cases.

O'Gorman [115] then points out that throughput here is the problem; somewhere this verification exception handling has to be done by someone. But the throughput problem is compounded if some form of biometric screening, say with face images, is performed. O'Gorman notes "Even if the probability of a false positive is set really low, 'to inconvenience as few as possible,' [at the expense of high probability of false negatives] there may be still be many false positives."

"Biometric Screening – Why does it point to me?"

Screening protocol: Match passenger face images with government face image database:

1. Assume a system that checks each person's face against a database of 25 alleged terrorists.

2. Use a best-case false positive (FP) rate for face of: 0.001 (= 0.1%).

3. If 300 people are requesting Jumbo Jet access
   7 of those will likely match suspected attackers
   ($25 \times 300 = 7500$ matches are performed, which gives $0.001 \times 7500 = 7$ false positives).
   That is, for each terrorist, 0.3 people on average will be ID'd $\times$ 25 terrorists = 7 people.

4. Again, the false positives have to be screened through some exception handling procedure (which is part of the screening protocol).

Note here that face screening is done at a False Negative (FN) rate of 10% to 20% on the ROC. This means that a person in database **N** (the most wanted) has a 80% to 90% chance of being identified; of course, this also means that a person in database **N** (the most wanted) has a 10% to 20% chance of *not* being identified. However, it is here where the deterrence factor comes in. Publication in the popular press of these type of error numbers may be hard to explain, since just like running a verification system at low FA (high security) one would desire to run a screening system at low FN (high security).

These two different authentication methods (Sections 7.2 and 7.2)run at opposite sides of the ROC, which may explain that vendors catering the criminal identification market could have difficulties in the access control (1 : 1) market.

O'Gorman then continues with the question: "How does 1 in 1000 rate result in 7 in 300 false positives?" The short reason is given in Point 3 above. (The reader may skip through to Expression 5.)

But, being precise, the false positive rate (FP) of screening one passenger against a screening database of n alleged terrorist depends on the database size n in the following way

$$\mathrm{FP}(n) = 1 - \prod_{i=1}^{n} (1 - \mathrm{FP}_i), \tag{2}$$

where the FP_i is the false positive rate of terrorist i. The FP_i are nonidentically but surely independently distributed random variables and therefore $\mathrm{FP}(n)$ is a random variable. Therefore we take the expectation (average) of $\mathrm{FP}(n)$ in (2) and obtain

$$\begin{aligned} \overline{\mathrm{FP}}(n) &= E\left[\mathrm{FP}(n)\right] = 1 - \prod_{i=1}^{n} \left[1 - E\left(\mathrm{FP}_i\right)\right] = 1 - \prod_{i=1}^{n} (1 - \overline{\mathrm{FP}_i}) \\ &= 1 - (1 - \overline{\mathrm{FP}})^n = 1 - [1 - \mathrm{FP}(1)]^n, \end{aligned} \tag{3}$$

where $E(\mathrm{FP}_i) = \overline{\mathrm{FP}_i} = \mathrm{FP}(1)$, the false positive rate when matching a passenger's face with *one* terrorist face. Expression (3) can be approximated for larger n and small $\mathrm{FP}(1)$ using the Taylor series expansion, as:

$$\mathrm{FP}(n) \approx n \times \mathrm{FP}(1). \tag{4}$$

Matching a data set $\mathbf{M}$ with m members, the population of a Jumbo jet $m = 300$ now requires m matches against a database $\mathbf{N}$ of n terrorists. Each passenger has to be matched against each terrorist, where the probability of a false positive is $\mathrm{FP}(n)$ as in (4). The number of false positives is then given by

$$\#\ \mathrm{FP\ for\ Jumbo\ jet} = m \times \mathrm{FP}(n) = m \times n \times \mathrm{FP}(1). \tag{5}$$

7.3 Face, finger and voice

Let us start with some illuminating numbers on error rates of biometric authentication systems. These number are directly taken from O'Gorman [114, 115], who in turn got then from [100, 124, 14], the most recent competitive and comparative testing of face, finger, and voice engines. O'Gorman gives the following table, which are the best error numbers found in [100, 124, 14]. All these competitions are technology evaluations (a). In this table, finger, face and voice are all operating in the order of 10% false reject; best

Authentication	**False reject**	False accept
Screening	False negative	**False positive**
Finger	3 to 7 in 100 (3-7%)	1 to 100 in 100K (0.001 - 0.1%)
Face	10 to 20 in 100 (10-20%)	100 to 10K in 100K (0.1-10%)
Voice	10 to 20 in 100 (10-20%)	2K to 5K in 100K (2-5%)

Table 8: Best and worst case error rates for face, voice and finger [115].

case false accept rates are then 1 in 100K ($10^{-5} = 0.001\%$) for finger, 100 in 100K ($10^{-3} = 0.1\%$) for face and 2K in 100K (2%) for voice, respectively.

7.4 Iris and hand

There are standard biometric evaluations, both technology and scenario evaluations procedures, proposed by Mansfield and Wayman [103] ("Best Practices in Testing and Reporting Performance of Biometric Devices"), maintained by National Physics Laboratory in the UK. The tests we found on iris and hand were done according to these recommendations.

Iris

The test scenario in [102] is verification within "a normal office environment" with cooperative, habituated users. The tests were conducted with 200 volunteers, over a three-month period with 1-2 months between enrollment and verification. The subjects were selected by invitation and, according to the study, few dropped out during enrollment. A test facility was used with the biometric authentication systems positioned roughly as in Figure 27b. (More than just iris and hand were tested in [102].)

	False reject	False accept	FTE	FTA
Iris	0.0%	2.0%	0.5%	0.0%
Explanation	Two different iris images are falsely matched	Two images of the same iris fail to match	Iris image cannot be acquired for enrollment	Iris image cannot be acquired for verification

Table 9: The iris verification error rates from [102].

The iris authentication error rates found in [102] are summarized in Table 9 for a factory selected operating point. The FTE (Failure to Enroll), FTA (Failure to Acquire), FA and FR are as defined as in Section 5. Also introduced in that section is the FTU (Failure to Use) rate, the probability that a person will not use a voluntary biometric authentication system and will stick to the legacy system. (Compare this to the old lady who insists on dealing with human tellers at the bank.) The volunteers were selected by invitation and most of them accepted the invitation, is unclear from [102] what the FTU is.

A test of iris identification that dates back to 1996 is the one from Boucher et al. [21]. In this test the iris authentication system was running not in *verification* mode but in *identification* mode. That is, the authentication protocol is just a single biometric with no claimed identity. Subjects just present the iris and the system makes the correct or incorrect decision, where an incorrect decision is either a FA or a FR. For such identification systems, the FA rate is a function of m, the number of subjects in database $\mathbf{M}$. Following (4) we have

$$\mathrm{FA}(m) \approx m \times \mathrm{FA}(1).$$

The FA rate is affected by m; the FR rate however is not.

In a first test by Boucher et al. [21], 895 iris identification transactions were recorded with 106 actual rejections, i.e., a FR rate of 11.8%. The authors observe that many people have problems enrolling with *one* of their eyes and can only enroll with their "dominant eye." These people have less chance of correct identification opposed to people enrolled with two eyes, who have two chances of correct identification. If users that were able to enroll with two eyes were allowed to do so, this FR rate therefore drops to 10%. These figures are contested by the iris identification system manufacturer (see attachment to [21]), nevertheless since so few independently estimated error rates for iris can be found, we adopt the 10% FR rate as worst case, though it is rather anecdotal. Neither the tests by Mansfield et al. nor the tests by Boucher et al. show any false accepts.

Wildes in [173] reports on an iris system developed at the Sarnoff Corporation that is not used in a commercial authentication system (yet). The iris database was collected using 60 irises from 40 people, that contain identical twins and a range of iris colors, blue, hazel, green, and brown. No false accepts and no false rejects were reported. However, as the author notes himself, the test is of no statistical significance because the data set is too small (Table 10). Recent testing performed by King [86] for DoD a test with 258 participants and 186,918 identification attempts, reporting 6% FR and 2 potential false accepts (FA $\approx$

0.001%). Because the data were not collected during testing, the two false accepts cannot be reproduced. Sixty people felt that the identification took too long, which is probably due to the repeated image sequence identification attempts. Using the 30-rule [122], to observe a 10^{-5} FA one needs to produce 30 errors (see Section 11.2). Hence, $30 \times 10^5 = 3\text{M}$ attempts are really needed, which implies that the 10^{-5} FA is therefore very anecdotal. It means however that the FA for iris is not zero.

	False reject	False accept
Iris	2 to 10 in 100 (2-10%)	$\geq 10^{-5}$ (0.001%) **??**

Table 10: Best and worst case error rates for iris from reports [21, 86, 102].

Why the error rates for iris are as reported is unclear. When running at such an excellent false accept rate $\text{FA} = \epsilon$ running at an equally excellent false accept rate $\text{FA} = 2\epsilon$ should result in lower false reject rates (than FR = 2% to 10%). This should be possible to do, no matter how the distance between two biometric sample is defined. That is, the operating point should really be shifted for this application.

Boucher et al. [21] gave a number of reasons for the false iris rejects (which may be or may not be stale right now):

1. User or environmental error: Presenting the wrong eye; not keeping the eye open; reflections from external light sources that obscure the iris.

2. Reflection from glasses: Glasses produce glare in the image. This could be avoided by users, except when the glasses are very thick.

3. User difficulty: Hair obscuring the iris image; difficulty focussing; problems with identifying based on non-dominant eye.

The error numbers (FA and FR) for iris are controversial. The Failure to Enroll (FTE) rate is perhaps the most controversial, there is anecdotal evidence that a segment of the population just cannot enroll with the iris biometric. Some systems require the subject to wiggle around until a series of lights appear to be aligned in order to get a good close up of their iris. Perhaps some of the FTE is due to this poor user interface. In general, obtaining a high-resolution image of an iris within possibly hostile ambient lighting conditions and with unpredictable specular reflection conditions is a nontrivial task. Computer vision experts are of the opinion that if the developers have solved this task, they have accomplished quite a feat. This is especially true if the imaging system works when the subject wears small, hard contact lenses that may even obscure the iris.

Hand geometry

An experimental hand geometry authentication system is found in [74], the only hand verification system published in the open literature. Here the hand geometry authentication system was trained and tested using a group of 50 users, resulting in 500 images; 140 images were discarded due to incorrect placement of the hand by the user. The authors note that this will create failure to enroll (FTE) problems and false rejects because of non-habituated users and conclude that therefore user adaptation of this biometric is necessary.

The system is tested by enrolling one person with two images and matching all other images to the enrolled hand geometry representation and a technology evaluation **a** of Figure 27 with results as in Table 11.

Note that four different operating points are reported in this tabale. Acceptable FR of 3% may be achieved but this is at a FA of 15%, which is not a reassuring number from a security point of view.

False accept rate	False reject rate
1 : 7 (15%)	1 : 33 (3%)
1 : 10 (10%)	1 : 20 (5%)
1 : 20 (5%)	1 : 10 (10%)
– (0%)	1 : 3 (30%)

Table 11: False reject rate, false accept rate pairs for the system in [74].

It should be mentioned that these are the numbers achieved by an experimental university prototype, and do not reflect the accuracy of commercially available systems. However, the error rate reported in [74] are the only estimates obtained by scenario **a**. Other error rate estimates for hand geometry authentication systems that are found in the literature are based on tests scenarios of type **b**. We look at two studies, a 1991 Sandia study [63] and the above mentioned Mansfield report [102].

The 1991 Sandia report "A Performance Evaluation of Biometric Identification Devices" by Holmes et al. [63] is in an "office environment" as in Figure 27. Quoting: "The verifier tests at Sandia were conducted in an office-like environment; volunteers were Sandia employees and contractors. A single laboratory room contained all of the verifiers [including the hand geometry system]. Each volunteer user was enrolled and trained on all verifiers." In an other part of the report we find: "Nearly 100 volunteers attempted many verifications on each machine."

However, in short, what the study finds is that the equal error rate, where the FA equals the FR, for the tested hand geometry verification system is approximately 0.2%.

Hand geometry is also tested using scenario **b** in [102] (as mentioned above, the tests were conducted with 200 volunteers, over a three-month period with 1-2 months between enrollment and verification). This report finds a $\mathrm{FTE} = \mathrm{FTA} = 0.0\%$; it further roughly seems to estimate the hand geometry EER at roughly 1%. Putting these numbers together, we get Table 12.

	False reject	False accept
Hand geometry	2 to 10 in 1000 (0.2-1.0%)	2 to 10 in 1000 (0.2-1.0%)

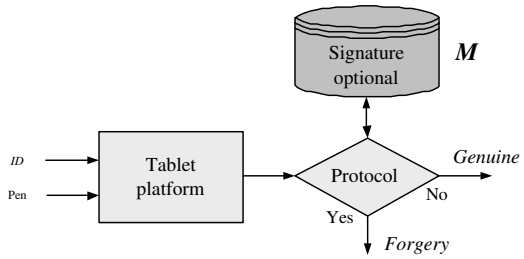
Table 12: Estimated error rates for hand geometry from reports [63, 102].

Sandia [63] also reports on a user survey regarding the verification systems that were tested. Questions like: *Which machine do you feel: rejects you the most often?* and *is most friendly/fun?* were asked. The answers for hand verification were remarkably favorable. The manufacturer subsequently used the number of positive responses and the number of negative responses to define an *acceptance ratio* for each biometric device. Hand geometry, based on that measure, is *more than 16 times* more accepted than the second runner-up, a retina verification system; this number is also quoted by the manufacturer [179] (of course).

7.5 Signature

Signature is commonly called a biometric but is misses one of the necessary conditions as prescribed in [31]; signature does not have the characteristic of *permanence* in that a person can change his or her signature

pretty much at will. One could, if one wishes, classify signature as a knowledge authentication method but we will continue to refer to signature as a biometric.



Signature verification

Figure 28: Signature verification and decisions.

Authentication decisions in general in terms of accept and reject are called *genuine* and *forgery*, respectively, see Figure 28. The signature verification area has over the years developed an interesting way of classifying the quality of a forgery as:

1. *Zero-effort forgery:* A zero-effort forgery is not really a forgery but a random scribble or signature of another individual.

   The FA and FR rates of a verification algorithm can be measured by just taking other signatures as forgeries, i.e., the same way as in test scenario **a** of Figure 27.

2. *Home-improved forgery:* These are forgeries which are made when the forger has a paper copy of the signature in possession and has ample time at home to practise and copy the signature at home. The imitation is based on just a static image of the signature.

3. *Over-the-shoulder forgery:* Here the forger is present while the genuine signature is written. The forger does not just learn the spatial image of the signature but also the dynamic properties by observing the signing process, which could be a curve in a five-dimensional space, see Section 4.6.

   Over-the-shoulder and home-improved are not distinguished when testing signature verification matchers. These are called amateur, semi-skilled or even skilled forgeries [120].

4. *Professional forgery:* These are forgeries produced by individuals that are skilled in the art of handwriting analysis. These people do not necessary have special skills in forging but they do know how to circumvent obvious problems and they can use their knowledge to produce high-quality forgeries.

This terminology and methodology probably traces back to the long history of handwriting authentication in the judicial system. The motive also has been the strong desire to detect forgeries right from the start of research in automated signature verification, simply because there have been many instances of the need for signature forgery detection and writer authentication over the ages.

In that sense, the signature verification literature makes an attempt to define levels of severeness of an attack on the verification system, i.e., impersonation attacks as we have called it in most other places in this document. That is, a distinction is made between *passive false accepts* and *active false accepts* (discussed in Section 12, as attacking the "lambs"). So, from the beginning the area defined different sets of error numbers associated with a signature verifier, loosely defined by the type of attack: attack by random signatures,

attacks by amateur forgers, and attacks by professional forgers, if you will. For none of the other biometrics has much effort been devoted to the skill level of the impersonator. In the voice verification area Doddington et al. [41] have made an initial attempt at qualifying the skill level of the attacker by introducing a category of speakers called *wolves,* those speakers that are particular good at imitating other voices (Section 12).

The signature research area selected more complicated ways of testing a verification system. Moreover, signature as a biometric has made rapid advances over the years, mainly perhaps because rapid advances on the input side see, e.g., [46]. These rapid developments probably impeded the development of a standardized testing and evaluation culture in this technology area. Quoting Dolfing from his PhD Thesis [42]: "In general, the comparison of signature verification approaches is difficult due to the different equipment, algorithms, models and databases. There is no standardized benchmark or database. Additionally, the type and quality of forgeries used to estimate the verification accuracy is different for each approach. This leads to false acceptance, false rejection and equal-error rates which make no sense without information about other system parameters."

Automated signature verification has a long history starting with approaches to static signature verification as described as early as 1977 [111]. An early survey paper from 1988 by Plamondon and Lorette [120] covers both static and dynamic approaches to signature verification. They note that signature outperforms initials and handwritten passwords. They further note dynamic signature is superior and that (in 1988) static writer verification (from a full page of text) performs about as well as static signature verification. Judging from [22] where it is concluded that signature and writing offer complementary information rather than specifically very correlated information, the issue whether writer verification is superior to static signature verification is still unresolved.

Unfortunately, the study [102] does not include a user study (scenario evaluation **b**) of dynamic signature; and we have to go back to the 1991 Sandia report [63] to find some numbers from a user study of a commercial dynamic signature verification system. Error numbers for the verification protocols one-try, two-try and three-try are given, which shows a large improvement from the one-try to the two-try numbers, see Table 13. This may indicate that a subject is much better able to put the signature so that it can be matched in the second try, perhaps because habituation of the biometric on that particular device is poor.

Dynamic signature		
	False reject	False accept
Three-try	2.06%	0.70%
Two-try	2.10%	0.58%
One-try	9.10%	0.43%

Table 13: The signature error numbers from [63] allowing one or more tries.

Dolfing in [42] does give comparative error estimates of a number of dynamic signature verification systems and we include in this list also the system described in [93]. The IBM system described by Worthington et al. [176] does best with $\mathrm{FR} = 1.77\%$ and $\mathrm{FA} = 0.28\%$ based on the first signature. Given that the proprietary database of signatures that is used to test the system of [176] probably consists of habituated signatures, it may be best to compare the two-try protocol of the system of Table 13 to the IBM system; for the system [93] error rates less than $\mathrm{FR} = 0.2\%$ and $\mathrm{FA} = 0.2\%$ This means that we find some consistency between signature accuracy numbers obtained by a technology evaluation (**a**) and scenario evaluations (**b**). We roughly combine the one-try numbers of both tests and arrive at Table 14.

The error rates for static signature seem to be about a factor 10 less accurate, in the range 2-5%.

	False reject	False accept
Signature	2 to 10 in 100	2 to 5 in 1000

Table 14: Best and worst case error rates for *dynamic signature* from [42, 63].

Interesting enough the matching technology that is used in [176], referred to as "regional correlation" and is more akin to approaches found in fingerprint and face representation technologies. It does better than speaker recognition based on Hidden Markov models.

Signature, as a man-made biometric, is doing surprisingly well in comparison to the other biometrics. The reported error rates that we find are consistently in the range of $\mathrm{EER} = 0.2\text{-}0.5\%$, however, the signature verification has different evaluation criteria in that error rates for different types of forgeries are measured. The error rates are measured for different types of forgeries, like *zero-effort* forgery, *home-improved* forgery, through, *professional forgery* and it even more difficult to relate the different studies, lt alone relate signature to the other biometrics.

7.6 Summary

Table 15 summarizes *the best error rates* as we found in the literature and as they are described in Sections 7.3 and 7.4. It should be emphasized that and it cannot be emphasized enough that face, finger and speech with test scenario **a**; iris and hand are tested with test scenario **b**; and, signature is tested with both scenarios, technology **a** scenario **b** evaluations (Figure 27).

	False reject / (FN)	False accept / (FP)	Evaluation type
Fingerprint	3-7 in 100 (3-7%)	1-100 in 100K (0.001-0.1%)	**a**
Face	10-20 in 100 (10-20%)	100-10K in 100K (0.1-10%)	**a**
Voice	10-20 in 100 (10-20%)	2K-5K in 100K (2-5%)	**a**
Iris	2-10 in 100 (2-10%)	$\geq 10^{-5}$ ($\geq 0.001\%$)	**b**
Hand	1-2 in 100 (1-2%)	10-20 in 1000 (1-2%)	**b**
Signature	10-20 in 100 (10-20%)	2-5 in 100 (2-5%)	**a** & **b**

Table 15: Roughly the error rates that can be found in the literature, based on scenario and technology evaluation.

The problem in general with any biometric authentication system, again, is of *volume*, especially if using one attempt identification protocols. The more matching attempts for verification the more false rejects. But what is worse, the higher the number of terrorist in the most-wanted list n subjects, the more false positives. This carries through to (positive) identification using with m members. In an identification application a false positive is like a false accept for security purposes. Hence, when security is involved, the immediate conclusion is that it is *never* a good idea to use "pure identification" because both practical and theoretical error rates of biometrics identification ($1 : m$ matching) are high, in which case the exception handling needs much attention [114].

However that when using any of the biometrics in Table 15 for screening (or negative identification), the only acceptable FP rate (if any of the FP rates are acceptable) is roughly running the authentication systems at the FN rates in the table. This means that fingerprint will than run at 3%-7% and face and voice run at

10%-20%. This means that a person in database **N**, the screening database, has anywhere from 3% to 20% chance at *not getting caught*.

Finally note that the tabulated values were measured at operating points that might not be suitable for a particular application (decreased FA can be traded for increased FR to some extent). As a rough rule of thumb, in practice an authentication system would be run with a FR around 5% (merely inconveniences the user), while a screening system would typically require an FR of 0.1% (subjects who require exception handling).

8 On the individuality of a biometric

There has been much interest in the *individuality* of the various biometric identifiers. Loosely speaking, this has to do with comparing a biometric with a password [134]; and it is related to how easy it is to randomly "guess" a biometric machine representation, given current sensing capabilities. The interest in the individuality of a biometric arose because of (at least) two reasons. The first reason is the work done by John Daugman on iris identification, see, for example [38]. In this work, Daugman proposed degrees of freedom of iris non-match score distribution as a measure of individuality, or uniqueness of the iris pattern. A second reason for the recent interest in individuality is due to many recent challenges to fingerprint, signature, and writer authentication expert testimony received in the courts. Here we will follow the landmark paper by Pankanti et al. [117]. It is a story about fingerprint, but similar situations for other biometrics will likely happen.

Fingerprint identification is routinely used in forensic laboratories and identification units around the world [92] and is accepted in courts of law for nearly a century [36]. Until recently, the testimony of latent fingerprint examiners was admitted in court without much scrutiny and challenge. However, in the 1993 case of Daubert vs. Merrell Dow Pharmaceuticals, Inc. [81], the Supreme Court ruled that the reliability of an expert scientific testimony must be established. Additionally, the court stated that when assessing reliability, the following five factors should be considered:

1. Whether the particular technique or methodology in question has been subject to a statistical hypothesis testing;
2. Whether its error rate has been established;
3. Whether the standards controlling the technique's operations exist and have been maintained;
4. Whether it has been peer reviewed, and published; and,
5. Whether it has a general widespread acceptance.

Subsequently, handwriting identification was challenged under *Daubert* (it was claimed that handwriting identification does not meet the scientific evidence criteria established in the Daubert case) in several cases between years 1995 and 2001 and several courts decided that handwriting identification does not meet the *Daubert* criteria. Fingerprint identification was first challenged by the defense lawyers under *Daubert* in the 1999 case of USA vs. Byron Mitchell [160] on the basis that the fundamental premises of fingerprint identification have not been objectively tested and its potential error rates are not known. The defense motion to exclude fingerprint evidence and testimony was first denied in January 2002 but was later (March 2002) allowed (for that particular case). Fingerprint identification has been challenged under *Daubert* in more than 15 court cases till date since the USA vs. Byron Mitchell case in 1999.

The two fundamental premises on which fingerprint identification is based are:

1. fingerprint details are permanent, and
2. fingerprints of an individual are unique.

The validity of the first premise has been established by empirical observations and is also based on the anatomy and morphogenesis of friction ridge skin. It is the second premise that is being challenged in recent court cases. The notion of fingerprint individuality has been widely accepted based on manual inspection (by experts) of millions of fingerprints. However, the underlying scientific basis of fingerprint individuality

has not been rigorously studied or tested. In March 2000, the U.S. Department of Justice admitted that no such testing has been done and acknowledged the need for such a study [159]. In response to this, the National Institute of Justice issued a formal solicitation for "Forensic Friction Ridge (Fingerprint) Examination Validation Studies" whose goal is to conduct "basic research to determine the scientific validity of individuality in friction ridge examination based on measurement of features, quantification, and statistical analysis" [159]. The two main topics of basic research under this solicitation include:

1. measure the amount of detail in a single fingerprint that is available for comparison, and
2. measure the amount of detail in correspondence between two fingerprints.

What does fingerprint individuality mean? As already indicated above, the fingerprint individuality problem can be formulated in many different ways, and yet another way is: given a sample fingerprint, determine the probability of finding a sufficiently similar fingerprint in a target population. Pankanti et al. [117] define this as the probability of a false association: given two fingerprints from two different fingers, determine the probability that they are "sufficiently" similar. If two fingerprints originating from two different fingers are examined at a very high level of detail, it may be found that the fingerprints are indeed different; in other words what is the theoretical lower bound on the false accept and false reject rates, often called the *intrinsic error rate.* Most experts and automatic fingerprint identification systems (AFIS) however declare that the fingerprints originate from the same source if they are "sufficiently" similar. How much similarity is enough depends on typical (within class) variations observed in the multiple impressions of a finger.

Pankanti et al. [117] subsequently developed a fingerprint individuality model that attempts to estimate the probability of a false association. They define *a priori* the representation of fingerprint (*pattern*) and the metric for the similarity. Fingerprints can be represented by a large number of features, including the overall ridge flow pattern, ridge frequency, location and position of singular points (core(s) and delta(s)), type, direction, and location of minutiae points, ridge counts between pairs of minutiae, and location of pores (see Figure 29). All these features contribute in establishing fingerprint individuality.

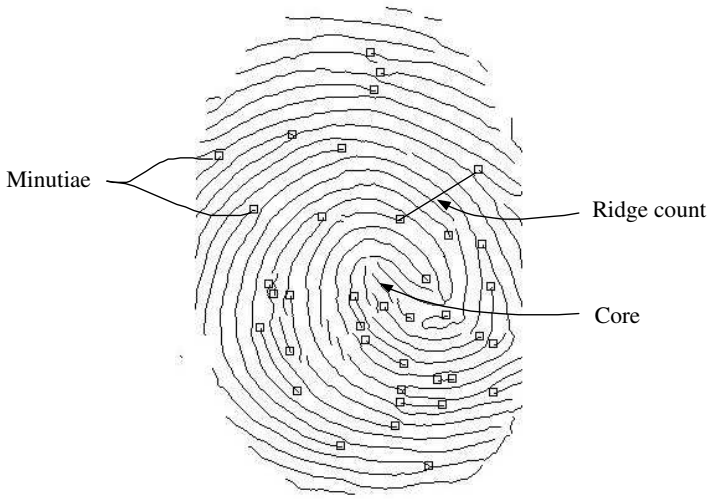


Figure 29: A fingerprint image of type "right loop" includes feature like the overall ridge structure, minutiae and ridge counts.

In [117] the authors chose a minutiae representation of the fingerprints because forensic experts use this type of representation, it has been demonstrated to be relatively stable, and it has been adopted by most of the automatic fingerprint matching systems. Given a representation scheme and a similarity metric, there are two approaches to determining the individuality of the fingerprints. A first approach is an empirical one (or technology evaluation) as discussed in Section 7. Here *representative* samples of fingerprints are collected and using a *typical* fingerprint matcher, the accuracy provides an indication of the uniqueness of the fingerprint with respect to this matcher. There are known problems (and costs) associated with collection of the *representative* samples, much of these issues are discussed in Sections 11.2 and 11.3 and Section 7. Therefore, a theoretical approach was selected in [117]: one models all realistic phenomenon affecting between class and within class pattern variations. Given the similarity metric they then theoretically estimate the probability of a false association.

The fingerprint individuality model by Pankanti et al. [117] estimates that matching two arbitrary prints (say, scanned at 500dpi and of size 0.75in $\times$ 0.6 in using minutiae representation is 10^{-22}).

Minutiae patterns are generated by the underlying fingerprints that are smoothly flowing oriented textures. Because minutiae positions are determined by the ridges, the minutiae points are not randomly distributed (see Figure 29). This typically implies that the probability of finding sufficiently similar prints from two different fingers is higher than that of finding sufficiently similar sets of random (directed) point patterns. By imposing fingerprint structural (e.g., ridge/valley position, ridge orientation) constraints on a random point configuration space to derive a more effective estimate of the probability of false association one can then enumerate all possible fingerprints, just like one can realistically estimate a password bit strength by using password banks and not using hard to remember random strings of 8 characters like 'rfghujjf.'

A similar enumeration scheme for fingerprints, yet simpler, is discussed in [134]. As mentioned previously, Daugman [38] studied iris individuality based on IrisCode representation (see Figure 30). IrisCode is a 256 byte binary code and the Hamming distance is used for comparing IrisCodes as the distance metric. Daugman assumed that chance match of independent bits of IrisCode follows binomial distribution and he then used empirical distribution of the non-match scores of IrisCode to estimate the number of independent bits in 256 byte IrisCode to be 173 bits.

Daugman's analysis estimates that the probability of finding two arbitrary IrisCodes sufficiently similar is 10^{-52}. This effectively implies that the probability of chance match of two unrelated iris codes to be 10^{-52}. The issues around these intrinsic error rates are very controversial, for example, there is much discussion about how precisely one can match such iris codes.

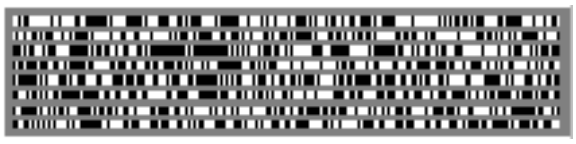


Figure 30: The bits strength of the iris representation is defined by using the Hamming distance between the codes.

We strongly feel that realistic error estimates of matching technology are those found in properly designed and executed technology evaluations (see Section 7). It is best to neglect all speculation about intrinsic error rates – error rates determined using statistically valid methods such as [117, 134] because these studies tend to be overly optimistic about the biometric individuality either due to unrealistic assumptions underlying the research or due to insufficient realistic data.

9 Application properties

There are plenty of places where biometric authentication can bring additional security. There are access control points of various security levels, and there are authorizations of all sorts of transactions. What is important in a biometric authentication application varies, among the salient features are:

1. what is the protected asset;
2. who are the authorized users and operators;
3. who are the adversaries of the application;
4. what are the consequences of security violations; and
5. what is the "cost" of inconvenience.

In this section we look at various applications from a point of view of these types of questions.

9.1 Wayman's application taxonomy

Wayman in [169] introduces an excellent taxonomy of biometric applications and it is best to simply quote Wayman verbatim:

> Each technology has strengths and (sometimes fatal) weaknesses depending upon the application in which it is used. Although each use of biometrics is clearly different, some striking similarities emerge when considering applications as a whole. All applications can be partitioned according to at least seven categories.
>
> *Cooperative versus non-cooperative:* The first partition is "cooperative/non-cooperative." This refers to the behavior of the "wolf," (bad guy or deceptive user). In applications verifying the positive claim of identity, such as access control, the deceptive user is cooperating with the system in the attempt to be recognized as someone s/he is not. This we call a "cooperative" application. In applications verifying a negative claim to identity, the bad guy is attempting to deceptively not cooperate with the system in an attempt not to be identified. This we call a "non-cooperative" application. Users in cooperative applications may be asked to identify themselves in some way, perhaps with a card or a PIN, thereby limiting the database search of stored templates to that of a single claimed identity. Users in non-cooperative applications cannot be relied on to identify themselves correctly, thereby requiring the search of a large portion of the database. Cooperative, but so-called "PIN-less," verification applications also require search of the entire database.
>
> *Overt versus covert:* The second partition is "overt/covert." If the user is aware that a biometric identifier is being measured, the use is overt. If unaware, the use is covert. Almost all conceivable access control and non-forensic applications are overt. Forensic applications can be covert. We could argue that this second partition dominates the first in that a wolf cannot cooperate or non-cooperate unless the application is overt.
>
> *Habituated versus non-habituated:* The third partition, "habituated/non-habituated," applies to the intended users of the application. Users presenting a biometric trait on a daily basis can be considered habituated after short period of time. Users who have not presented the trait

recently can be considered "non-habituated." A more precise definition will be possible after we have better information relating system performance to frequency of use for a wide population over a wide field of devices. If all the intended users are "habituated", the application is considered a "habituated" application. If all the intended users are "non-habituated", the application is considered "non-habituated." In general, all applications will be "non-habituated" during the first week of operation, and can have a mixture of habituated and non-habituated users at any time thereafter. Access control to a secure work area is generally "habituated." Access control to a sporting event is generally "non-habituated."

Attended versus non-attended: A fourth partition is "attended/ unattended," and refers to whether the use of the biometric device during operation will be observed and guided by system management. Non-cooperative applications will generally require supervised operation, while cooperative operation may or may not. Nearly all systems supervise the enrollment process, although some do not [171].

Standard environment: A fifth partition is "standard/non-standard operating environment." If the application will take place indoors at standard temperature (20^o C), pressure (1 atm.), and other environmental conditions, particularly where lighting conditions can be controlled, it is considered a "standard environment" application. Outdoor systems, and perhaps some unusual indoor systems, are considered "non-standard environment" applications.

Public versus private: A sixth partition is "public/private"." Will the users of the system be customers of the system management (public) or employees (private)? Clearly attitudes toward usage of the devices, which will directly effect performance, vary depending upon the relationship between the end-users and system management.

Open versus closed: A seventh partition is "open/closed". Will the system be required, now or in the future, to exchange data with other biometric systems run by other management? For instance, some State social service agencies want to be able to exchange biometric information with other States. If a system is to be open, data collection, compression and format standards are required.

This list is open, meaning that additional partitions might also be appropriate. We could also argue that not all possible partition permutations are equally likely or even permissible.

9.2 Weighting the factors

Any application can now be classified according to the seven above categories. Table 16 looks at various applications and weighs the importance of the various aspects of the different biometrics (as discussed in, e.g., Section 10). Therefore negative aspects of biometrics are listed in the left column, while the different applications are listed along the top row of the table. One could create a similar table listing the severity of the negative aspects for each specific biometric. Then, by conceptually multiplying the two tables together, once could derive application-specific fitness scores for each biometric.

*** include companion table here also? ***

The applications roughly divide up into three groups with different requirements:

- access control systems;

- screening applications; and,
- transaction authorization.

We cover these in the next three sections.

IMPORTANCE WEIGHTING	**Physical access**	**Stock trading**	**Airport access**	**Credit card authorization**	**ATM transaction**
Intrinsic properties:					
Required cooperation	Low	Low	High	Low	Medium
Social stigma	Medium	Low	High	Medium	Medium
Intrusiveness	Medium	Low	Medium	High	Medium
Missing population	Low	Low	Medium	Medium	High
Sampling properties:					
Inconvenience	Medium	Low	Medium	High	High
Required proximity	Low	Low	High	Medium	Medium
Acquisition time	High	Medium	High	Medium	Medium
Failure to enroll	Medium	Low	Medium	High	High
Failure to acquire	Medium	Medium	High	High	High
1 : 1 Matching properties:					
Template Size (bytes)	Low	Low	Medium	High	High
FR @ 10^{-4} FA (per 10M)	Medium	Low	Medium	Low	Low
FR @ 10^{-6} FA (per 10M)	Low	Low	Low	High	High
FA @ 10^{-1} FR (per 10M)	Medium	Medium	High	Medium	Medium
FA @ 10^{-3} FR (per 10M)	High	High	Medium	Low	Low
Discrimination (per 10M)	Low	Low	Medium	High	High
Technology properties:					
Installation cost	Medium	Low	Medium	High	High
Continual run cost	Medium	Low	Medium	High	High
Per-match cost	Low	Low	Medium	Medium	Medium

Table 16: Importance weightings for various applications.

Physical access and stock trading

These are the applications with typically small numbers of users enrolled in the application. The monetary consequences of authentication failures in terms of damages and losses can be easily in the order of one million dollars, or so. The user of the application has every reason to cooperate with the application because this type of authentication is a job requirement. Hence, this authentication application is not voluntary and there exists a relatively static access control list. These are typically the applications where the negative convenience aspects of any of the biometrics are weighed relatively low.

Airport screening

This application is in a class of its own because the consequences of authentication errors can be the direct loss of life. This is also an application for which the user community is composed of all sorts of constituents.

This in itself makes the design of biometrics authentication hard and costly. However here the higher costs of installing and running the installation is of medium importance, as opposed to the more specialized application such as stock trading where the costs are of lower importance.

Credit card transactions and ATM transactions

This application scenario is largely the opposite of physical control and stock trading. Credit card transactions and ATM transactions are applications with typically a larger number of users enrolled in the application and the monetary consequences of authentication failures in terms of damages and losses are only in the order of $500 or so. The user of the application does not have many reasons to cooperate with the application because the legacy solution may be just fine with the user. Moreover, because of lack of liability of the user for ATM fraud and credit card fraud (currently – this may change), there is even less incentive to use and cooperate with the biometric authentication. These are typically applications where all the negative convenience aspects of any of the biometrics are weighed heavily because the requirements are high. Moreover, the cost of the installation has to be low because the applications may potentially be low-margin businesses. Therefore, the higher cost of biometric authentication for these applications are important negative decision factors.

9.3 Affordability and cost

From a cost point of view, again a very first issue in biometric authentication systems is the issue *verification* versus *identification*. Identification involves $1:n$ matching of biometric representations while verification involves $1:1$ matching. This means that identification involves by necessity large centralized databases, more secure communication infrastructure and in general n times more computational power than is involved for verification solutions.

Further, upfront investment is of course affected much by the properties of the individual biometrics as discussed in Section 10. That section looks at issues like maturity, scalability, sensor cost and size, and template sizes of the various biometrics. Upfront investment is further directly influenced by the choice between single or multiple biometric authentication protocols.

We mention in this section the costs that are added to authentication because of introducing biometrics in the authentication protocol. A large portion of the upfront investment for biometric authentication is system training, because in reality as we discussed in Section 11.1 the system will have to be trained during operation.

- *System training:* The improving, refining and tuning of the biometric matcher while the system is installed. This training task almost always has to be performed during operation of the authentication system (Section 11.1) at the site of operation and therefore could be expensive, but is needed to reach the required false accept and false reject rates.

  Reaching these specifications may further involve quality control of biometric acquisition, which could be application dependent. This may affect the convenience of the overall solution and negatively affect the throughput, thereby increasing costs.

  System training a system is related to enrollment. Enrolling a subject is the process of training a system with a biometric representation of the subject.

Besides this, there are a number of critical continuing cost factors:

- *Enrollment:* As noted this is an often underemphasized task. When security is the prime concern of an application, the enrollment should be more secure or at least as secure as the authentication. As seen in Section 12, enrollment is a process that has to be administered very carefully and therefore can be of significant cost. In effect, the enrollment has to be performed in a more secure and trusted environment than the authentication itself.

  If the enrolled population of a biometric application reaches a steady state, the enrollment costs may level off. However as discussed in Section 12, maintaining the enrollment database or access control and its security, in terms of data integrity may be a difficult task because of all the subtleties of an enrolled population in terms of Doddington's zoo [41].

- *User education:* A certain level of understanding of a biometric application is required from the user. Misunderstandings in the use of the application can have disastrous acceptance or throughput consequences. Marketing cost are a significant portion.

- *Supervisory labor:* Depending on the various job levels among the inspectors and operators of a biometric authentication system, certain skill levels in biometric applications are required.

  Many biometric applications are supervised by an inspector. These inspectors typically control the quality of the presentation (acquisition) of the biometrics; perform exception handling in the case of, say, *Failure to Acquire* (FTA, see Section 5). But in addition, very importantly, these operators may need to perform *visual matches* of the biometrics. (See Section 3.6 for a discussion of the contributions of operator errors to the overall error rates.)

- *Maintenance labor:* Human labor is needed to maintain the application. This labor ranges from highly skilled computer operators to less skilled personnel to perform mundane but important tasks as cleaning the biometric scanning devices.

- *Failure to enroll (FTE):* The FTE rate is something that is hard to estimate prior to system development and installation. FTE means exception handling and can be a significant extra hidden cost factor.

  When a biometric authentication system operates at high levels of FTE rates, conversion of the system from voluntary to mandatory may be very costly.

10 Selecting a biometric

We already discussed the error numbers for the various biometrics that can be found in the literature in Section 7. In this section we mention other properties that are important for our particular biometrics; while in Section 9 we discuss properties that are more related to the particular application. In Section 8, we discuss the intrinsic error rates of our six biometrics.

We look at issues like maturity, scalability, sensor cost and size but first we seek to answer the often asked question, what is the size of a biometric machine representation *or template* of a biometric? This becomes important when biometric templates need to be stored on smart cards or mag stripes.

10.1 Storage issues

When looking at biometrics from a storage point of view, a first issue in biometric authentication systems is *authentication* versus *identification*. Identification involves $1:n$ matching of biometric representations while authentication only involves $1:1$ matching. This means that identification involves large centralized databases and that in general n times more infrastructure is needed. Sometimes the storage requirements depend almost exponentially on n (for example, [50]). These are geometric hashing schemes that deal with increasing n by using increasingly more storage, or rather a large machine representation of the entire database. These are approaches that basically are space-time tradeoffs. Note here, incidentally since as is found in Section 8, the realistically estimated bit strengths of many biometrics are rather low and hence the theoretical (or empirical) estimates are. This means that with the growth of computing power, algorithms such as in [50] are becoming obsolete because soon biometric database searches will be possible within reasonable times using sequential search, which is always more accurate than hashing or tree search algorithms. That is, the scalability of a biometric is determined by its intrinsic error rates and not by the available compute power or other resources.

Storage may also become an issue when the algorithms use the notion of cohorts (see Section 11.1.1) because then if the individual templates are stored in a distributed database of biometric templates, for each individual biometric representation somehow the cohort or the world (if this is the cohort model) needs to be included.

While these are issues to be aware of, here we only report template sizes for the biometrics of interest in this document. A template is a machine representation of a biometric sample. Hence a template in some way describes the acquired biometric sample so that automated matching of biometrics can be done as precisely as possible. Template sizes for the various biometrics vary a lot, from very small for iris to larger for voice. Ruggles in a study [141] prepared for the California State Legislature reports (in bytes) iris $= 256$, finger $= 512 - 1K$, hand geometry 9 bytes. Using Ruggles, and roughly following the advances in the state of the art, we summarize storage requirements as in Table 17.

	Fingerprint	Face	Voice	Iris	Hand	Signature
Template size	$<$ 200 bytes	$<$ 2K bytes	$<$ 2K bytes	256 bytes	$<$ 1K bytes	$<$ 200 bytes

Table 17: The machine representation of biometrics vary in size.

Template sizes are not an indication of individuality, bit strength, or false accept rates of a biometric. For fingerprints, maybe, the bit strength is probably somewhat related to template size. For other biometrics, in general, there is no direct relation between any other property of a biometric and the template size.

As noted before, for privacy or security reasons, it may be desirable to store the authentication credentials, including the biometric sample or machine representation, on a distributed database, such as a collection of smart cards in the possession of the enrolled population. Otherwise there has to be some kind of trust between the users of an application and the database owner and operators.

In this version of the document, we do not consider processing times, since these are much harder to pin down because they depend on how efficiently the code was written and what sort of processor it runs on. Moreover, compute time is starting to be a non-issue as processors get ever faster (how long on a 6 GHz machine?).

10.2 Other characteristics

Other aspects of the six biometrics are considered in Table 18. These are six biometric properties, some of which are desirable for one type of application, as discussed in Section 9. Three of the properties are directly related to the user interface. Another one, maturity, is also related to the user interface assuming that the user interface improves over the years.

	Fingerprint	Face	Voice	Iris	Hand	Signature
Sensor type	contact	unobtrusive	unobtrusive	unobtrusive	contact	contact
Sensor size	small	small	very small	medium	large	medium
Sensor cost	$< \$200$	$< \$50$	$< \$5$	$< \$3000$	$<\$500$	$< \$300$
Maturity	very high	medium	medium	medium	high	medium
Scalability	high +	medium	low	very high	low	high −

Table 18: Comparison of six popular biometrics.

In any application, and particular for biometrics applications, possibilities for error are mainly at the user interface. These include user actions, presumably mostly because of visual information supplied to the user. These actions may be wrong simply because of severe usability problems with authentication interfaces. Often, the user interface is all too clumsy and incomprehensible and usability studies are needed. These are applications for real people, one cannot expect users to understand that they have to push the '**Start**' button in order to shut down the machine (quoting Gerstner from some interview).

The issue *scalability* in Table 18 is a little different. The table expresses what we found in Section 7. Face and fingerprint, and also the other biometrics, are maybe much less accurate than commonly believed, yet still the accuracy numbers of the table indicate that there is a group of strong biometrics (iris and finger) and a set of weak biometrics (hand, face, voice). It is important to understand that there is breakthrough technology needed to elevate the weak biometrics into the class of strong biometrics, i.e., simple incremental improvements and tuning are not sufficient. The rate of improvement is too slow for practical solutions for many authentication protocols to be possible yet.

For civilian applications, where security is not an issue, the additional convenience (no password) may justify the use of biometrics. These type of identification systems may run at unexpectedly high error rates though, see Section 10. Conversational biometrics, as discussed in Section 14.3 is a solution to biometric authentication by incorporating knowledge as credentials, including the communication of the possessions such as account numbers. A similar approach can be used with other biometrics, also. The knowledge K in this authentication protocol is of private and personal matter and typically only known by the subject. But just like the answer to the question *What is your mother's maiden name?* correct answers to questions do not

decrease formally the FA and FR rates of the process over authentication using the raw rates for speech. It is possible that an imposter has all this personal and private information about a person, but it is not probable. Error rates for conversational biometrics are therefore hard to estimate formally.

10.3 Positives and negatives of the biometrics

A technology and application exploration of the area of biometrics can be found in [16]. This report looks at today's existing installations and more application aspects for each biometric. A number of positives and negatives for each of our biometrics are enumerated. It is useful to repeat and expand upon this list here:

Fingerprint

pros:

- There is a long tradition of the use of fingerprint as immutable identification in law enforcement. Though there are some recent challenges of that premise (see the section on the "Individuality of a biometric," Section 8, which tells the judicial story behind fingerprints).
- There exist large legacy databases of fingerprints, albeit largely of criminals. However, the California, Colorado, Florida, and Texas Departments of Motor Vehicles are working to establish a fingerprint biometric system for drivers licenses and record data.
- Fingerprint lends itself well to forensic investigation, i.e., the study of latent fingerprints [51]. Criminals often leave a trail of fingerprints (e.g., hotel, car, door knob, glass, weapon) that allows for reconstruction of facts after the events take place.
- A fingerprint is easily sampled using low tech means, and the size and price of fingerprint readers are still declining. The conversion of fingerprints into digital images is getting easier, better and cheaper. There are low cost fingerprint scanners available (under $100) that are already in widespread use in many access control applications.

cons:

- Fingerprints suffer from poor public acceptance in some countries because of the very strong relationship between fingerprint and criminal history. This has its advantages but does not help in the acceptance of finger as a biometric.
- There is a large variation of the quality of the fingerprint over the population. The appearance of a person's print depends on age, grease, cut or bruised fingers; i.e., on occupation and life style in general.
- Sampling an image of a fingerprint is a matter of pressing the finger against the platen of a fingerprint reader. This creates technical problems as discussed in Section 4, and may also create problems with cleanliness of the sensor and public hygiene.

Face

pros:

- Photos of faces are widely used in passports and drivers licenses where the possession authentication protocol is augmented with a photo for manual inspection purposes, therefore there is wide public acceptance already.
- Face recognition systems are the least intrusive from a biometric sampling point of view, and is completely without contact.

- The biometrics works, or at least works in theory, with legacy photograph databases, videotape, or other image sources. Face recognition can be used for screening for unwanted individuals in a crowd, in real time
- Good for verification.

cons:
- A face needs to be well lighted by controlled light sources in automated face authentication systems. This is only a first challenge in a long list of technical challenges associated with robust face authentication.
- Face currently is a poor biometric to be used in a pure identification protocol (see Section 7.2). It performs better in verification, but not at accuracy rates that are sometimes claimed.
- An obvious circumvention method is disguising, which will easily cause false negatives in screening applications; i.e., the undesirable but disguised person is not identified.
- There is some criminal association with face identifiers since this biometric has long been used by law enforcement agencies (mugshots).

Voice

pros
- Like face, voice is a natural biometric under certain circumstances (phone) and like face sampling, voice sampling is quite unobtrusive
- There exist great public acceptance of this biometric partly because of its naturalness and partly because there is no criminal association with voice.
- The voice biometric only requires inexpensive hardware and communication. Voice is therefore very suitable for pervasive security management.
- This biometric allows incremental authentication protocols. For example, the protocol prescribes to wait for more voice data when a higher degree of recognition confidence is needed.
- The voice biometric may achieve high accuracy and flexibility when combined with knowledge verification, in the authentication protocol called *conversational biometrics* (see Section 14.3).
- The voice biometric allows for checking identity continuously (Section 3.9), i.e., in a passive way the voice can be authenticated at regular intervals.

cons
- Imitation by skilled impersonators may be possible; this potential susceptibility has not been studied in any detail yet.
- With the text to speech technology improving, e.g., [150], it becomes possible to create nonexistent identities with machine voices; this is of course only possible when enrollment is also remote. But this brings us to deeper questions not quite in the scope of this document.
- There simply is not as much redundance in the audio signals as there is in visual signals. Consequently, the results of audio processing for semantic understanding and biometric authentication is very dependent on the ambient sounds.

Iris

pros:
- Iris is claimed and believed to be the most accurate, especially when it comes to FA. Iris has very few false accepts (the important security aspect), with only anecdotal evidence. Therefore iris is a good biometric for pure identification applications.

- Given that the iris sample acquisition process is solved using unobtrusive distant cameras, the sensing of the biometric is without physical contact
- Iris receives little negative press and there may therefore be an accepted biometric, the fact that there is no criminal association (yet) is helping here.
- The dominant commercial vendor claims that iris does not come with much training costs.

cons:
- Though iris is a good biometric for identification, large-scale deployment is not possible because there are so few iris databases. (Iris is currently more appropriate for verification.)
- The iris is a tiny bit of human appearance and therefore sampling the iris pattern requires expensive input devices.
- The performance of iris authentication may be impaired with tinted glasses, sunglasses and contact lenses; people may have to take them off or out.
- The implementation of the iris sampling system has a small field of view; the input device may need to be tilted for short or tall subjects.
- The iris biometric, in general, is not left as evidence on the scene of the crime, which makes it a questionable biometric for criminal applications.
- From the iris image, certain deceases (e.g., diabetic retinopathy) can be diagnosed and hence sampling the iris may have unexpected privacy consequences.

Hand

pros:
- There exist definite public acceptance for the hand biometric because it is already used at Disney World, INSPASS (INS, Passenger Accelerated Service System) and at various universities for verification in, e.g., the University of Georgia.
- There exists at least one scenario evaluation of hand geometry as a biometric, which shows that hand is a good biometric for verification. (There is not much known about error rates for hand as an identifier.)
- It is claimed that hand geometry measuring is an easy do-it-yourself operation.

cons:
- As face, hand geometry is not a very distinctive from one person to the next. Therefore, as discussed in Section 7.2, this biometric is a poor one to select for pure identification.
- There exists no legacy national database of hand geometry measurements, which may impede progress with this biometrics. The INSPASS database is of unknown size.
- As with fingerprint, hand geometry is measured when a subject presses the biometric against a platen (although there are ways around this). Such contact may be cause for some public hygiene concerns.

Signature

pros:
- Signature is a man-made biometric where forgery has been studied extensively, therefore forgery is detected even when the forger managed to get a copy of the authentic signature or even more characteristics of a signature.

- At enrollment stage there is already some possibility of detecting forged signature. This can be used to detect the impersonation (forgery) of existing identities (signatures) at enrollment time.
- Training is intuitive and fast, and people understand (from speech recognition technology) that the system needs to be trained. They intuitively understand what it means to enroll in a way not causing false accepts on your signature.
- Signature verification in general has fast response (about 1 sec per signature on the "old" 486DX-33 computer) and low storage requirements.
- A signature verification is independent of the native language of the user.
- Because signature is a man-made biometric and you can use any kind of information as your signature: name, second name, or even nice curves (e.g. the artist formerly known as Prince).
- Very high compression rates do not affect shape of the signature (100-150 bytes).

cons:
- Although there is much precedence for using signature to authenticate documents, this may give the perception that signature is not secure enough for protecting airports, etc.
- The area of study of the individuality of signature, is a little different in that it is rich in ideas and has a much better understanding of personalization (forgery) efforts. Unfortunately, the community is not very disciplined in their system testing culture.
- A 5-dimensional pen (see Section 4) may be needed to arrive at the desired accuracy. The 5-dimensional pen as input device records pressure and angles too, this makes the hardware costly. It is unclear how well the X-Y algorithms that need no pressure information perform. That is, the effectiveness of signature for access control using existing tables, is unclear.

11 Training and testing the system

There are many things peculiar to biometric authentication systems. For one, the algorithms for automated biometric matching have to be trained, and therefore tested. Biometric system training really has two aspects, a first one is the training that is done during enrollment (as discussed in Section 12) where a system is trained to authenticate a particular biometric instance (i.e., a person), or rather samples from that instance. A second process where system training takes place is is during the design and tuning of biometric authentication systems.

11.1 System training

For a given biometric solution or application it is hard to gather training data that is a proper representation of the target population, the population of individuals that will eventually use the system. Therefore it may be hard, or impossible, to accurately determine the probabilities of match for any given input and reference sample before an installation is in place. This means it is hard to properly set the system variables beforehand. Often, the only choice is to design and develop the system with some test sample set and then tune the system *after* it is installed. When training the system, there should be sufficient training data [19, 20] available to accurately estimate these match probabilities. (Of course, the notion of training data is unknown in legacy password authentication systems.)

To understand how a biometric authentication application is trained, it should be defined exactly how the biometric authentication decision (ultimately) is made. *How well* these decisions are made can then only be tested, i.e., it can be tested if a biometric application operates at expected error rate levels (application parameters, Section 5). There are many statistical pitfalls when measuring these operating parameters, somehow statistically estimating biometric operating parameters may lead to statements that are of questionable origin. All this is discussed in the next section.

Automatic biometric authentication, at a minimal, has to be able to *compare* an acquired biometric to a stored biometric, the template. This is an issue of confirming whether an acquired biometric sample is in some sense "close" to a template and, therefore, belongs to the same individual. Hence any automated biometric application has to be able to decide the $1:1$ matching problem and compute probabilities of match (and ideally a probability of a *mismatch*). Biometric identification involves $1:n$ matching, where the identity of a subject is recognized by comparing the biometric sample to that of a database of n enrolled individuals. Hence, for identification systems, no unique identifiers are required *per se*. What is additionally required for biometric identification is a (central) database DB for storing the templates.

Biometric applications seldom store a sample of a biometric identifier (a biometric signal), rather a compact representation (template) of the biometric sample is computed and stored. Either way, both biometric samples and their representations/templates are patterns, i.e., the pattern $P = S(\mathcal{B})$. The pattern is a sample of biometric $\mathcal{B}$, the individual. Authenticating a person with a biometric sample is then formulated in terms of hypothesis testing. Let the stored biometric sample or template be pattern $P' = S(\mathcal{B}')$ and the acquired one be pattern $P = S(\mathcal{B})$. Then, in terms of hypothesis testing, we have

$$\begin{aligned} H_0 &: \quad \mathcal{B} = \mathcal{B}', \quad \text{the claimed identity is correct} \\ H_1 &: \quad \mathcal{B} \neq \mathcal{B}', \quad \text{the claimed identity is } \textit{not} \text{ correct.} \end{aligned} \tag{6}$$

When making such a decision, two types of errors can be made:

FA: The probability of deciding H_0 is true while in reality H_1 is not true; the probability of erroneously granting a subject access to the application (*False Accept*).

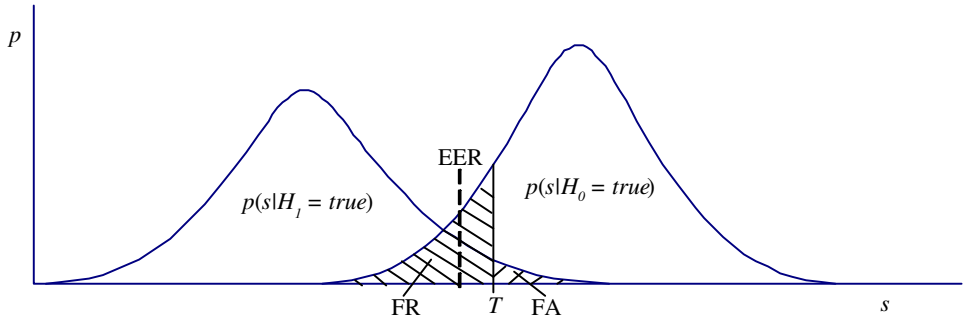


Figure 31: Probability density distributions of mismatch (intruder) and match (genuine) scores. The hatched areas denote FR and FA.

FR: The probability of deciding H_1 is true while in reality H_0 is true; the probability of erroneously denying a subject access to the application (*False Reject*).

For the moment assume that this decision (6) is based on a degree of match s, often referred to as a *score*. If s is the score of a *mated* pair, s will tend to be high, when s is the score of a mismatching biometric, s will be low. This is shown in the graphs of Figure 31 with the density of match scores $p(s/H_0 = true)$ on the right and the density mismatch scores on the left $p(s/H_1 = true)$. Decisions then are based on a *decision threshold* T and if for some subject (identity), $s > T$, the subject is granted access.

The Equal Error Rate are the error probabilities where we have

$$\text{ERR} = \text{FA} = \text{FR}.$$

This error rate occurs at the threshold T_{EER} where the dashed areas under the curves $P(s/H_1 = true)$ and $P(s/H_0 = true)$ equal, that is, where H_1 is true but the subject is accepted (False Accept) and where H_0 is true but the subject is rejected (False Reject).

In Figure 31, the threshold T is selected higher than the Equal Error Rate (EER). This means that the biometric matcher is optimized for low FA rate, or equivalently, higher security (than at the EER). The choice of T is part of satisfying the design parameters as discussed in Section 5. In general, *training* of an application amounts to deciding T so that the authentication operates at the required FA and FR. Databases are available publicly (for example, for fingerprint authentication, from a recent Fingerprint Verification Contest [100], more databases are given in Section 15).

We can cast the $1:n$ identification problem as a hypothesis testing problem, for example, as a *screening application.* For this, let us define DB to be the database of (voluntary / involuntary) enrolled individuals, then

$$\begin{aligned} H_0: &\quad \mathcal{B} \in DB, \quad \text{the individual is in the database.} \\ H_1: &\quad \mathcal{B} \notin DB, \quad \text{the individual is } \textit{not} \text{ in the database.} \end{aligned} \tag{7}$$

The database DB could, for instance, be the set of individuals on some "most-wanted" list. The database DB could be a collection of such criminal databases.

The errors, using Figure 21, then are

FN: The probability of deciding H_1 is true while in reality H_0 is true; the probability of erroneously missing a subject that should be "flagged" (*False Negative*).

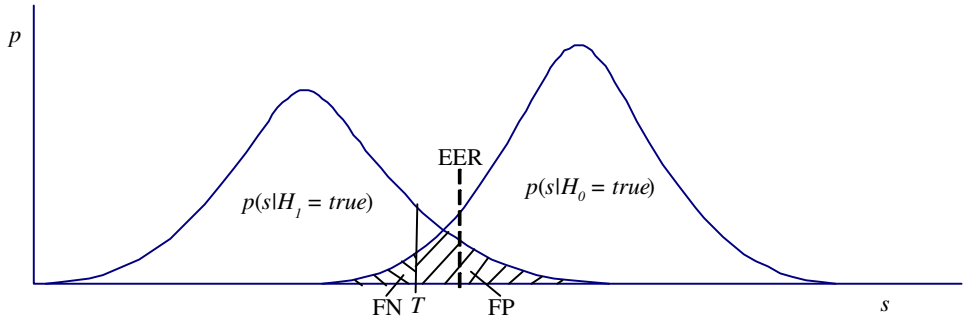


Figure 32: Probability density distributions of mismatch (unenrolled) and match (enrolled) scores. The hatched areas denote FN and FP.

FP: The probability of deciding H_0 is true while in reality H_1 is not true; the probability of erroneously "flagging" an innocent subject (*False Positive*).

For identification purposes, the decision threshold T in Figure 32 will be less than the Equal Error Rate (EER). This means that the biometric matcher is optimized for low FN rate, or equivalently, high probability of finding matching biometric identifiers in database DB. This in turn means higher FP rate leading to possibly large lists of matching database entries for each biometrics query. There are two modes that a biometric identification system can be designed for:

1. *Screening mode:* As above, each subject is biometrically matched against a database DB to determine if the subject bears biometric similarities (or other similarities) to subjects in DB. The FN should be low, in order *not* to miss suspects; the FP should be low in order *not* to generate too many suspects for each enquiry because all suspect will have to subjected to further scrutiny.

   Training a biometric identification application to operate in screening mode means mainly designing the application such that *not all subjects* are biometrically matched. The biometrics in such applications are those collected in today's criminal databases. The challenge of such an application is *when* to select subjects for biometric matching; and, *vice versa*, how to select a small set of "most-wanted" individuals.

   Fundamental limitations of particular biometrics limit the size n of the biometric database that can be used. Logistical application limitations determine for each biometric the throughput that can be expected. These are two of the biometrics properties discussed in Section 10.

2. *Identification mode:* A subject is biometrically matched against database DB to determine if the subject biometric identifier is represented in database DB. The FN should be low so that few genuine subjects are not recognized; the FP should be low so that no intruders are accidentally recognized.

   Training a biometric identification application to operate in identification mode means designing the application such that any genuine subject is recognized and any imposter is not mistaken for some other identity in the database. This is quite a difficult design objective and is described, e.g., in [50].

   Any biometrics has intrinsic error rates or an associated bit strength that determines the scalability of the biometrics. There is a great variation of scalability between the various biometrics.

As noted before, the only reason to operate biometrics in identification mode is to offer convenience, Section 6.2. There is never too much security, and therefore biometric identification is not appropriate when

security is a prime concern, since it is very hard to train these systems to operate at low FP and FN at the same time, Section 6.1.

11.1.1 Cohorts

An interesting idea, "anti speaker modeling" or "cohorts" comes from the voice recognition community [139], see also [91]. Voice verification methods do not only use a model describing who the speaker is (biometric machine representation), but also a model describing all other speakers. The reason for this has been that the probability of a match (a likelihood statistic given the input data) cannot be evaluated directly in many of the Hidden Markov model voice recognition algorithms and the likelihood of observing the voice data itself needs to be determined. Determining this likelihood of the voice data by modeling all possible speakers $\mathcal{W}$ is of course not possible. Basically two techniques exist to approximate this likelihood [138]. One technique (cohort modeling), estimates the set by a finite set $\mathcal{C}_i$, which is representative for the group of speakers which resemble speaker i, called the set of cohorts of speaker i. The other technique (world modeling) reduces the sets $\mathcal{C}_i$ to a set of size 1, containing only one fictitious modal speaker whose model is trained on a pool of data from many different speakers, who represent the "world" $\mathcal{W}$ of possible speakers.

No conclusive difference in performance between the two techniques in text dependent speaker verification has been found, but world modeling has some computational and algorithmic advantages over cohort modeling:

- Much less storage is required for the model parameters.
- It is computationally more efficient during recognition, since only one anti-speaker likelihood needs to be evaluated,
- It does not require the selection of cohort models $\mathcal{C}_i$ per speaker i.

The world model approach is the commonly adopted one. This world model is computed in different ways:

- A completely speaker-independent implementation.
- Gender- or handset-dependent implementations, where the choice for the world model depends on the gender of the claimed identity or the type of handset used by the actual speaker [135].

A promising avenue of research is cohort modeling for other biometrics. Similar such techniques are of course already in use, for example, for training neural networks by giving negative examples. The above mentioned cohort modeling is within a framework of probabilistic modeling.

11.1.2 Protocol

As discussed in Section 3.1, authentication is achieved through some protocol $R(P, K, B)$ an operator R that is a function of the credentials P, K, and B of a particular subject. Biometric protocols, as we have seen already, are plenty, a few examples:

- "Three tries and you're out."
- Use either left index finger or right index finger.
- Use either iris or finger.

It is not always obvious that each individual biometric should be operating at its individual optimal decision point, as seen in Section 14.2, and care should be taken to train a biometric authentication system for different authentication protocols. That is, protocol: "use either left index finger or right index finger" and 'use both left index finger and right index finger" may have to operate at different decision thresholds T.

11.2 System testing

As will be further described in Section 7 (our investigation of the reported error rates for the various biometrics), there broadly exist two types of biometric authentication system testing:

Technology evaluations: Biometric sample databases are collected off-line and the system is tested on these databases.

Scenario evaluations: A group of volunteer subjects is drafted, and this group uses the system over a time period while statistics are collected.

Both methods are used by the biometric industry but it seems that academia tends to use (often public) databases, i.e., technology evaluations. Of course, during scenario evaluation, databases of biometric samples are processed (and maybe collected) and the estimation of error rates becomes a matter of computing estimates based on sample data as introduced in Section 11.2.2.

11.2.1 Scenario evaluations

Often used biometric test procedures are described in the report "Best Practices in Testing and Reporting Performance of Biometric Devices" by Mansfield and Wayman [103]. Quoting the report:

> "Our aims are:
>
> 1. To provide a framework for developing and fully describing test protocols.
> 2. To help avoiding systematic bias due to incorrect data collection or analytic procedures in evaluations.
> 3. To help testers achieve the best possible estimate of field performance while expanding the minimum of effort in conducting their evaluation.
> 4. To improve understanding of the limits of applicability of test results and test methods."

This report covers both user tests (called scenario evaluation) and testing on databases (called technology evaluations). A host of recommendations is given in this report, ranging from how best to select a crew of volunteers to recommendations on what fingerprint sensors to use (sensors that comply with certain standards and technical specifications as described in [34]). For volunteer selection, which is perhaps the most challenging aspect of biometric testing, the report notes that volunteer selection may bias the tests. The authors note that individuals with objections against the technology and individuals with physical challenges may be less likely to participate in the testing. It then may be necessary to unevenly select subjects from the volunteers to make the crew as representative as possible and to make sure that problem cases are not under-represented. Unfortunately, however, the authors do not give many practical recommendations on this issue. They note: "Our understanding of the demographic factors affecting biometric system performance is so poor, that target population approximation will always be a major problem limiting the predictive value of our tests."

The report *Best Practices in Testing and Reporting Performance of Biometric* [103] covers scenario evaluations and technology evaluations. Scenario evaluation test procedures described in this report were applied in the testing of seven biometric systems in [102]; the results obtained in the latter report are used extensively in Section 7, where we discuss the biometric error rates reported in the literature.

The report further notes that it is important to relate the amount of data that is used to the quality of the error estimates. That is, given the data and the computed error estimate, how confident can one be in these estimates; or in other words, what are the confidence intervals of our estimates. It is known that there exists dependence among the match scores of sequences of mated biometrics, which makes confidence interval computation difficult. In [20] we introduce the "subsets bootstrap" to compute confidence intervals. This method is based on the Bootstrap (see [48]), motivated by the "moving blocks" bootstrap [95] and explained in the next section.

11.2.2 Technology evaluations

Such testing is basically an issue of defining the performance parameters that need to be measured, defining how these parameters can be estimated from test data, and deciding which databases to use for performance evaluation. Suppose as in Figure 33 the evaluation uses databases Z_1 and Z_2 of biometric samples (and databases A and B for system training); for each database it is known which images are associated with which individuals. Hence, by selecting pairs of samples from the databases Z one can select a number of matching biometric sample pairs (mated pairs), measured from the same subject, and a number of pairs of biometric samples that do not match, measured from different subjects. Upon matching all these pairs, one can obtain M match scores and N mismatch scores; we denote these sets as $\mathbf{X} = \{X_1, X_2, ..., X_M\}$ and $\mathbf{Y} = \{Y_1, Y_2, ..., Y_N\}$. Hence we have M match scores and N mismatch scores, where $M \ll N$ because there are much fewer mated pairs than mismatched pairs. Note that the mismatch pairs are sample biometrics of random individuals and that the type of forgeries are therefore zero effort forgeries.

Figure 33: One common test scenario is based on databases.

Given the set of match scores $\mathbf{X}^a$ from a first matcher, matcher a, and a set of match scores $\mathbf{X}^b$ from a second matcher b, can one use match score sets $\mathbf{X}^a$ and $\mathbf{X}^b$ to make statistical valid statements about the matchers and statistically valid statements comparing matchers a and b? After score normalization, one can estimate all kinds of parameters of the probability distributions F^a and F^b from the sets of samples $\mathbf{X}^a$ and $\mathbf{X}^b$.

For example, one can estimate the sample means $\overline{X}^a$ and $\overline{X}^b$ of data sets $\mathbf{X}^a$ and $\mathbf{X}^b$

$$\overline{X}^a = \frac{1}{M}\sum_{i=1}^{M} X_i^a \text{ and } \overline{X}^b = \frac{1}{M}\sum_{i=1}^{M} X_i^b.$$

Now we may want to conclude that matcher a is worse than matcher b when $\overline{X}^a < \overline{X}^b$; after all this lower average match score implies that scores of mated pairs for matcher a are on average lower then those

for matcher b. The question however is whether we have enough data to accurately estimate $\overline{X}^a$ and $\overline{X}^b$ or is the fact that $\overline{X}^a < \overline{X}^b$ something that happened by chance? In other words, we want to be able to say how confident we are in the fact that hypothesis $\overline{X}^a < \overline{X}^b$ is true in reality, i.e., $E[X^a] < [X^b]$. This is an issue of statistical hypotheses testing and we need to be able to compute confidence intervals for $\overline{X}^a$ and $\overline{X}^b$. This is shown in Figure 34: If we have little data, the confidence intervals are large and they overlap and we cannot conclusively say that matcher a is worse than matcher b. With more data, we can come to this conclusion confidently because the confidence intervals *do not* overlap.

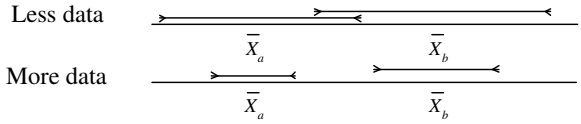


Figure 34: Hypotheses testing is a matter of computing confidence intervals.

Let us drop the superscript and show how the bootstrap [48] can be used to compute a confidence region around estimate $\overline{X}$ using the set of samples $\mathbf{X} = \{X_1, X_2, ..., X_M\}$. Note that the samples in set $\mathbf{X}$ are the only data we have. The bootstrap proceeds, therefore, by assuming that the empirical distribution of this data $\mathbf{X}$ is a good enough approximation of the true distribution F. The bootstrap then prescribes sampling, with replacement, the set $\mathbf{X}$ many (B) times:

1. Calculate the estimate $\overline{X}$ from the sample $\mathbf{X}$. This is the estimate around which we want to establish a confidence interval.
2. *Resampling.* Create a bootstrap sample $\mathbf{X}^\star = \{X_1^\star, ..., X_M^\star\}$ by sampling $\mathbf{X}$ with replacement. Here in set $\mathbf{X}^\star$, X_i may be represented multiple times and X_j may not be represented at all.
3. *Bootstrap estimate.* Calculate $\overline{X}^\star$ from $\mathbf{X}^\star$.
4. *Repetition.* Repeat steps 2-3 B times (B large), resulting in $\overline{X}_1^\star$, $\overline{X}_2^\star$ ... $\overline{X}_B^\star$.

One can compute (say) a 90% confidence interval by counting the bottom 5% and top 5% of the estimates $\overline{X}_i^\star$ $i = 1, ..., B$ and subtracting these estimates from the total set, as illustrated in Figure 35. The leftover set determines the interval of confidence. What one will find however is that this bootstrap confidence interval is much smaller than it should be because the bootstrap is only valid for identically, independently distributed (i.i.d.) random variables. There exists dependence among the match scores $X_1, X_2, ..., X_N$ because match scores from the same finger are dependent. However, the way the bootstrap sets $\mathbf{X}^\star$ are obtained from the original set of match scores $\mathbf{X}$ *does not* replicate the dependence among the X_i. Therefore the bootstrap estimates $\overline{X}_i^\star$ have lower variance than would be the case if the proper bootstrap sets were sampled from $\mathbf{X}$.

Therefore, in [20] the *subsets bootstrap* motivated by the moving-blocks bootstrap [95] is proposed. Assume that the set $\mathbf{X}$ with M match scores is obtained from D different biometric instances (e.g., from D fingers). Then the set can be regrouped as $\mathbf{X} = \{\mathcal{X}_1, \mathcal{X}_2, ..., \mathcal{X}_D\}$. The subsets bootstrap now resamples with replacement, using the sets $\mathcal{X}_i$, i.e.:

1. Calculate the estimate $\overline{X}$ from the sample $\mathbf{X}$. This, of course, is the same estimate as above.
2. *Resampling.* Create a bootstrap sample $\mathbf{X}^\star = \{\mathcal{X}_1^\star, ..., \mathcal{X}_D^\star\}$ by sampling $\mathbf{X}$ with replacement. Here set $\mathbf{X}^\star$ may contain X_i multiple times and may not contain X_j at all.

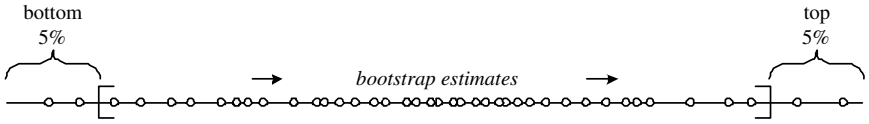


Figure 35: Obtaining a bootstrap estimate amounts to ordering and counting .

3. *Bootstrap estimate.* Calculate $\overline{X}^{\star}$ from $\mathbf{X}^{\star}$.

4. *Repetition.* Repeat steps 2-3 B times (B large), resulting in $\overline{X}_1^{\star}$, $\overline{X}_2^{\star}$ ... $\overline{X}_B^{\star}$.

The confidence intervals are then obtained the same as above (Figure 35).

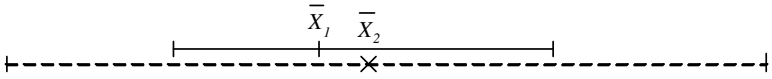


Figure 36: Estimates computed using increasingly more samples should result in properly included confidence intervals.

Figure 36 shows two estimates of $\overline{X}_1$, and $\overline{X}_2$, where $\overline{X}_1$ is obtained from a set of samples $\mathbf{X}_1$ and $\overline{X}_2$ is obtained from a set of samples $\mathbf{X}_2$ and $\mathbf{X}_1 \subset \mathbf{X}_2$. The subsets bootstrap guarantees this type proper inclusion of better confidence intervals as shown in Figure 36. Figure 37 shows empirical distributions of two sets of match scores, $3,600$ and $9,000$ match scores computed by using a fingerprint matcher. Confidence intervals around the more precise empirical distribution estimate are contained in the confidence intervals of the less precise empirical distribution.

11.3 Comparing matchers using a test data set

Assume that it is possible to define some measure of accuracy. A statistical estimate of this accuracy measure can be obtained, which is computed from a test database. Comparing two matchers is then a process of subtracting two accuracy estimates. The difference between two accuracy estimates is itself an estimate and can only be determined within some confidence intervals, e.g., with 90% confidence Matcher A is better than Matcher B.

Three things are needed for matcher comparison:

A data set: A total number of D fingers are used and d ($d > 3$) impressions per finger are acquired. The number of impressions per finger is the same for each finger to avoid that one, or a small number, of fingers dominates the estimates.

We have $M = D\,d\,(d-1)/2$ match scores (symmetric matcher) and $N = d^2\,D\,(D-1)/2$ mismatch scores. The number D then is determined by the "30-error rule," [122] if one needs to measure error rates of one in X or $(1/X), 30X$ samples are needed. That is, to measure a FR in the order of 10^{-4}, some 300,000 mismatch scores are needed.

An accuracy measure: Given the data set, accuracy can either be expressed as one error rate pair (FA, FR) at a particular operating point on the ROC curve (a plot of multiple error rate pairs) or as the ROC itself. The only way to measure these accuracies is by statistical estimation. Comparing two matchers is the comparison of two error rate pairs estimates or the comparison of two estimated ROC curves.

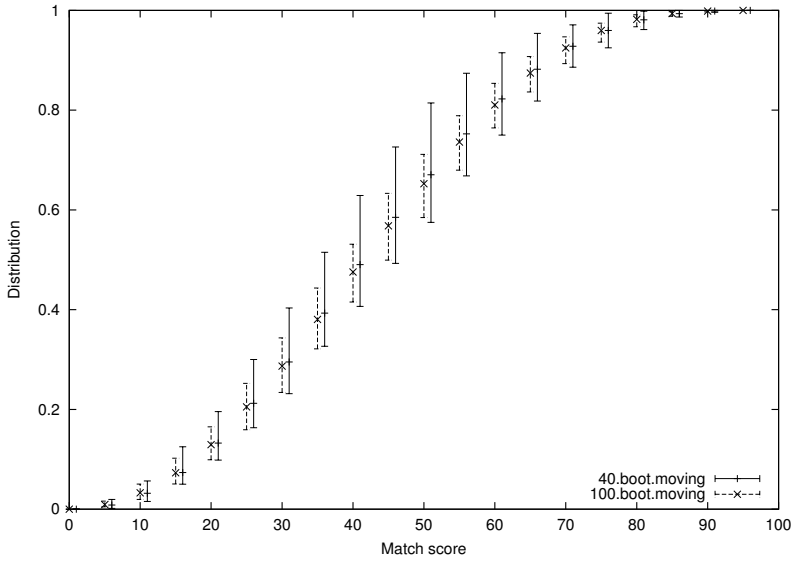


Figure 37: The confidence intervals computed using the larger number of match scores lie within the confidence intervals computed using the smaller number of match scores.

The former is achieved by formal statistical tests. The latter is probably best achieved visually, if one wants to compare ROC curves.

Confidence intervals: Confidence intervals are often computed under the assumption that the M match and N mismatch scores are independent.

This independence assumption *results in severely underestimated confidence regions*. Confidence intervals on the (FA, FR) estimates are to be determined using statistical bootstrap techniques that make no assumptions about the underlying error distributions.

With this:

- One can answer hypotheses, such as: "Is the FA of matcher A the same as the FA of matcher B?" Here a degree of confidence in the answer can be determined.
- One can determine an area of confidence around a ROC curve. This allows for overall visual comparison of matcher A to matcher B and gives an idea of the statistical significance of the comparison.

Yet, still, it has to be recognized that the answer is only true for a particular application in as far as the data set represents the target population.

Using the ROC curve

It may be impossible to determine if matcher A is better than matcher B by somehow quantitatively or qualitatively comparing the two associated ROC curves. As shown in Figure 38, a first matcher (matcher A) does not necessarily have to be consistently better than a second matcher (matcher B), at every operating point. If one defines the accuracy at some operating point as FA + FR, it is seen that at operating point X, matcher A is more accurate; while at operating point Y, the reverse is true (matcher B is more accurate). If

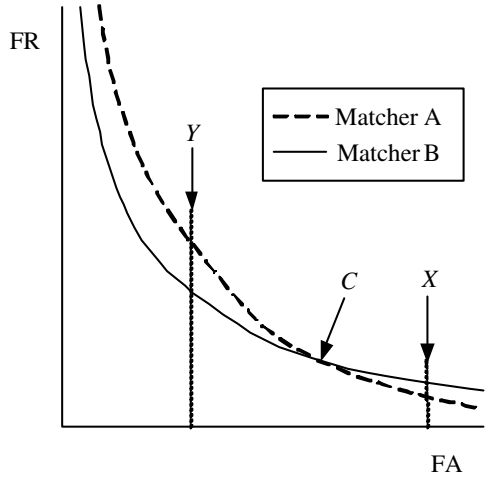


Figure 38: We are interested in operating points, not in ROC curves.

a low FA is more important, it is best to use matcher A at operating point X. If a low FR is the objective (at the expense of the FA), using matcher B at operating point Y is the optimal choice.

Therefore, it appears impossible to make any statements about differences in accuracy of two matchers based on just comparing two ROC curves. Only statements based on differences of two pairs of (FA, FR) estimates can be made.

Specific statistical test

The objective is to determine matcher A's accuracy is equal to matcher B's accuracy versus is matcher A somehow better than matcher B (matcher $A >$ matcher B). This will be estimated at a fixed operating point, $(\mathrm{FA}, \mathrm{FR}) = (Y, \mathrm{FR}$, say, with FA $= Y = 10^{-4}$, see Figure 38.

In terms of hypothesis testing, we have

$$\begin{aligned} H_0: &\quad \text{matcher } A \;=\; \text{matcher } B, \quad \text{the matchers are equal in performace.} \\ H_1: &\quad \text{matcher } A \;>\; \text{matcher } B, \quad \text{matcher a is better than matcher b,} \end{aligned}$$

which translates into

$$\begin{aligned} H_0: \mathrm{FR} \text{ of matcher } A &\;=\; \mathrm{FR} \text{ of matcher } B \\ H_1: \mathrm{FR} \text{ of matcher } A &\;>\; \mathrm{FR} \text{ of matcher } B. \end{aligned}$$

Such a question phrased in terms of hypotheses testing can be answered with any desired probability of being true. It can be statistically decided that $\mathrm{FR}\ \text{matcher}\ A \;=\; \mathrm{FR}\ \text{matcher}\ B$ with $x\%$ confidence for any given x.

12 Enrollment and database management

Enrollment is the process of selecting trusted individuals through some enrollment policy E_m and somehow storing machine representations of the m enrolled members in the verification database **M**. Enrollment is also the process of determining the undesirable individuals through some enrollment policy E_n and registering machine representations of the n selected individuals in the screening database **N**.

These processes are shown in Figure 39. Here **W** is the world population with, in a sense, a certain "ground truth" is available in the form of legacy databases, both civilian and criminal. The content of the legacy databases may or may not be an accurate reflection of the ground truth; after all, countries could have very liberal and corrupt passport procedures and other bureaucratic processes.

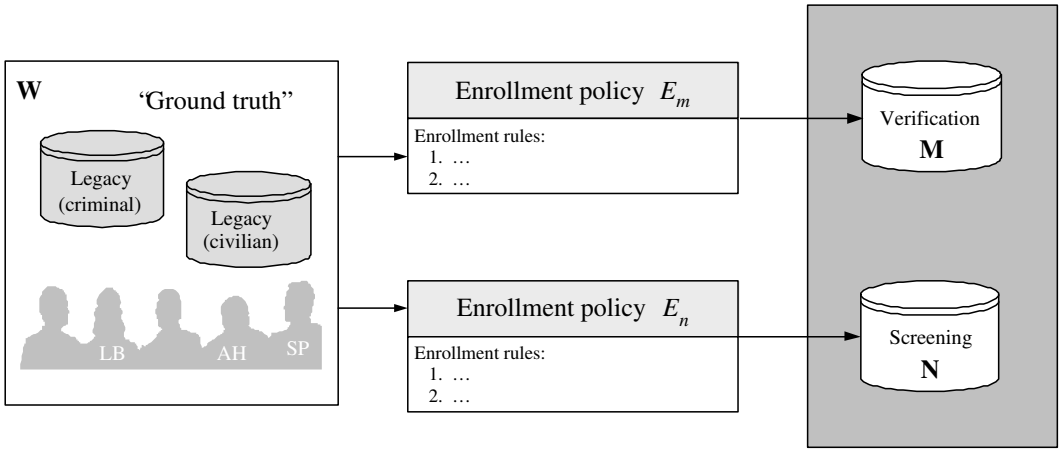


Figure 39: Enrollment consists of two processes: Building a member database **M** and (optionally) building a screening database **N** of undesirable people.

12.1 Enrollment policies

Enrollment policies are in a way very similar to authentication protocol. The difference is that an enrollment policy contains authentication protocols, and not vice versa. We define the enrollment policies E_m and E_n as:

E_m : Membership enrollment, this is a process of the registration of m trusted subjects $d_i, i = 1, ..., m$ in some database **M**. The enrollment could be based on an enrolled population for some application such as frequent flyers. Then of course the integrity of the database **M** is no better than the original frequent flyer database. In any case, a policy for enrolling subjects into **M** needs to be developed and published and for enrollment a subject needs to be authenticated by one or more authentication protocols. These protocols may have human inspectors in the loop.

Part of the membership enrollment policy most likely is that the trusted subjects d_i are inspected against the screening database, which is collected according to screening policy E_n.

E_n : This is a process of registration of n questionable subjects $d_j, j = 1, ..., n$ by storing machine descriptions of these subjects in some database **N**. The policies here are about when a subject d_j in a legacy

database is included in the screening database **N** and what information about a subject d_j is stored along.

The purpose of this database **N** is of dual use. It may be used as part of the enrollment policies, i.e., each trusted subject d_i needs to be matched against database **N**. It also of course is used for screening subjects d_k who are not a member of **M** when they request access.

Careful enrollment policies and quality control of the enrollment are central to the ultimate success of biometric authentication. Much care is needed during enrollment and the reason for this is explained in Section 12.3.

For enrollment process E_m, we repeat some of the things we already said in the introduction and expand on it a little. Enrollment policies that determine who are the eligible subjects **M** need to be defined for each application. These enrollment policies E_m include things like this:

- The "true" identity of a subject has to be (manually) verified or established somehow. What proofs of identity are accepted at enroll time?

- A definition of membership requirements for the prospective members $d_i, i = 1, ..., m$ needs to be defined. That is, it needs to be defined who the customer is.

  Ironically, this is sometimes called the *target population,* an unfortunate choice of words because this population *is also* the target population of the attacker.

- What credentials are issued and what credentials are shared with the application (biometrics, personal knowledge) for subsequent authentication. Choices made here will limit the ultimate flexibility of the authentication protocols as we saw in Section 3.

Again, during enrollment, other databases may be available, indicated by the databases *legacy civilian* and *legacy criminal* in Figure 39. Then, during enrollment a subject can be screened against the databases to determine whether a person has some undesirable credentials; these "credentials" can be civilian, i.e., credit report records or drivers license records; these credentials can be criminal, like previous arrests.

Enrollment policies that determine who are the undesirable subjects $d_j, j = 1, ..., n$ to be registered in screening database **N** need to be defined for each application. These policies includes things like:

- It needs to be exactly specified and agreed upon what type of information about a subject d documented in legacy databases, such, as FBI, INS, DMV excludes subject d_j from database **M**.

  The details of the policies for running such enquiries of legacy databases may or may not be publicly made available. Nevertheless, the enrollment policy somehow has to define this.

- It needs to be exactly specified and agreed upon what type of information about a subject d_j documented in legacy databases, such, as FBI, INS, DMV is that results in the inclusion of d_j in **N**.

Enrollment policies can only be developed through an open national and international debate. Consequently, this debate will probably be chaotic, like everything else in the biometrics area. The resulting enrollment policies will be also, like everything else, are a complicated matter of compromises.

12.2 Probabilistic enrollment

The enrollment process has as ultimate goal the assignment of a '1' or '0' to each subject $d_k, k = 1, ...$ in the world population **W**, meaning, either you are trusted or you are not. This is an unrealistic goal, and it is best to assign some probability p_i to each trusted subject $d_i, i = 1, ..., m$.

After all, after enrollment real world subject s_j is just a machine representation or token T_i, i.e., (ID_i, Pw_i, B_i), an identifier along with a password (Pw_i) and a biometric template B_i. And there is only a probabilistic link between this token T_i and real world subject d_i. It is therefore desirable that the enrollment process of Figure 39 also computes and stores likelihoods

$$p_i(ID = d_i/B_i), i = 1, ..., m.$$

That is, the likelihood of d_i given the stored template; or in a sense this expresses how well a subject's biometric matches his or her template.

Equivalently, one could have an expression for

$$p_i(ID = d_j; j \neq i/B_i), i = 1, ..., m,$$

which forms the basis of the cohort techniques (Section 11.1.1).

So, we only have some probabilistic association between a biometric template T_i and the real subject d_i, because after enrollment, a subject is *only* a machine representation (some biometric template and some ID) of the real-world biometric, which is the subject itself. As described in [17], this allows for better authentication decision-making and it allows the association of meaningful measures of the cleanness and security of the database **M** and therefore the security (integrity) of the installation.

12.3 The zoo

This brings us to the enrollment database population as a system variable, because we want to express the integrity of the database **M** somehow. Here we do not refer to the integrity of the credentials $T_i = (d_i, Pw, B_i)$, $i = 1, ..., m$. These credentials could be flat-out wrong to begin with because of circumvention, coercion and impersonation (see Section 13), and that means that because of sloppy enrollment the database is not clean anymore.

But even if the enrollment database were "clean" by some definition, an enrolled database of biometric identifiers brings with it, because of the fuzzy or probabilistic matching, the notion of subjects with strong and subjects with weak biometric identifiers ; that is, for the same biometric, one subject may be easy to authenticate because he/she has a very distinctive pattern while another subject may be hard to authenticate because his/her biometric pattern is not very distinctive. The speech recognition community realized early that there are subjects whose speech is easy to recognize and there are subjects that are hard to recognize, these subjects are called sheep and goats, respectively.

Doddington et al. in [41] introduced this notion of animals ("Doddington's zoo") to the area of speaker recognition and added some animals (see below). Classifying the enrolled subject in terms of animals is a good tool to understand the enrolled population of a particular installation. This classification is intended for voice recognition and for speaker identification, but is applicable to other biometric identifiers as well. It also amplifies the importance of careful enrollment, because if enrollment is neglected it can really become "a zoo out there" to quote Doddington et al.

The *sheep* and *goat* are the traditional categories:

Sheep: This is the group of subjects that dominate the population and authentication systems perform reasonably well for them simply because their real-world biometric is very distinctive and stable.

Goats: The group of subjects that are particularly difficult to authenticate with a poor real world biometric that is not distinctive, maybe damaged and variable over time.

This is the portion of the population that generates the majority of false rejects.

Doddington et al. added two more subject categories:

Lambs: These are the *enrolled* subjects who are easy to imitate. A randomly selected speaker from the general population is highly likely to be erroneously authenticated as an enrolled lamb. Equivalently, in the signature area, a *zero-effort forgery,* (in essence a scribble) is highly likely to be erroneously authenticated as an enrolled lamb. Simply because lambs have easy-to-forge signatures.

Lambs are the subjects, when enrolled in a database that constitute the portion of the enrolled population that are the cause of a false accept because they are eaten by wolves.

Wolves: These are subjects that are particularly good at imitating, impersonating or forging a particular biometric. That is, for example in speaker recognition their speech is likely to be accepted as another enrolled speaker's speech and these subjects will make successful intruders causing another type of false accept. Equivalently, in signature a professional forger is a wolf and if the forger has knowledge about signatures of enrolled lambs, the easiest thing to do for the intruder is simply to forge the signature of a known lamb.

A possible way to distinguish these two types of false accepts is to think of lambs causing "passive false" accepts and wolves causing "active false" accepts. This notion of the two types of false accepts is introduced in [18]. The passive false accept rate is measured by randomly matching random biometric samples to enrolled samples. However, each enrolled subject $d_i, i = 1, ..., m$ has its own false accept and false reject rate and hence attacking an installation by impersonating the right lamb will result in much higher error rates. However, actively attacking a system with unenrolled wolves will result in a significantly much higher (active) false accept rate. The classes are summarized in Table 19.

Doddington's zoo [41]	
Sheep	Well behaved subjects in the population.
Goats	The subjects that cause the false rejects.
Wolves	The subjects that attack other subjects in the population, causing active FA.
Lambs	The subjects that are attacked by other subjects, causing passive FA.
Chameleons	The subjects that both attack and are being attacked.

Table 19: Categories of subjects based on their impact on authentication system performance.

The *chameleons* were introduced in [18] because a matcher can be designed to be symmetric, i.e., distance template$_a$ to template$_b$ equals the distance from template$_b$ to template$_a$. We arrive at the strange situation that wolves $\equiv$ lambs.

Chameleons: These are the subjects who are both easy to imitate and are good at imitating others. They are a source of passive false accepts when enrolled and of active false accepts when being authenticated.

The classification in terms of animals depends on the way the match score probability is defined [18] and is still poorly understood. Classification of fingerprints in terms of their matching behavior is attempted in [19].

Hence, every biometrics implementation has its own zoo, which is composed of population groups in terms of the above animals. The exact composition of the target population in term of these animals greatly influences the accuracy of a particular application. Conversely, setting the Failure to Enroll higher (by, e.g., more quality control, see below) will result in the disqualification of the undesirable animals and the enrolled population will consist, as desired, of mostly sheep.

12.4 Biometric sample quality control

Many *random* false rejects/accepts are because of adverse and hostile signal acquisition situations. Poor input control perhaps constitutes the single most important reason for high false reject/accept rates. Apart from better user interfaces, there are two solutions to this. One can either probabilistically model and weigh all the adverse situations into the feature extraction/matching system or one can try to dynamically and interactively obtain a desirable input sample The latter is only possible in interactive overt authentication protocols involving cooperative users.

Automatic implementation of either strategy needs an algorithmic characterization of what a *desirable* pattern sample is, some quality measure. The term "quality" is then somehow related to "processability" of the sample. A system faces difficulty in analyzing poor quality samples and performs well when presented with good quality samples. It is important to quantify the concept of quality so that consistent actions can be taken for different types of undesirability, so that an appropriate corrective action can be suggested or taken by the system (e.g., apply enhancement) or by the subject (e.g., present the biometrics in a different, "better" way). Finally, it is desired that assessment of quality is fast and does not involve actually matching the given sample either with its mates or with its non-mates. Often, one single best sample is desired.

Conveniently handling biometric samples of diverse quality is important to any practical biometric authentication system. However, in theory, for almost all applications it is possible to compromise some convenience (ease-of-use or FR) by accepting only a certain quality of input. Not surprisingly, almost all operational biometric systems have an implicit or explicit strategy for handling samples of poor quality. Some of the simplest measures for quality control include the provision for presenting multiple samples to the system. In other schemes, the system provides a user with live visual display of the biometric that has been sampled (which is of course not practical for all biometrics).

This input quality control will result in higher FTE rates, accepting low-quality biometrics samples lowers FTE. Quality control of the input samples (especially during enrollment) is important to maintain a certain level of quality, i.e., with higher FTE, only good biometric samples will join the enrolled population. This is important because the number of users that, in one way or the other, are subject to some exception handling (e.g., a human gate agent) can be an added cost in a particular biometrics installation. Unfortunately, these subjects are hard to recognize thereby increasing the FR but perhaps more importantly, these subjects will also increase the FA rate. Hence, the FTE rate is an extremely important variable of an application because if it is too high an application simply becomes too costly. In some sense then, the optimal FTE for a biometrics application is a tradeoff between upfront investment and the continuing operating cost of a biometrics installation, i.e., $FTE \Rightarrow \mathrm{Cost}$, the FTE of an installation greatly affects the operating cost

13 Points of attack

Automated biometrics helps to alleviate the problems associated with existing methods of user authentication. Biometrics can improve convenience or security, or ideally both. Though, security weak points will exist or will be introduced in any biometric installation. These weak points will likely be found during the operation of a system, because the system will probably be attacked at these points.

Unlike password systems, which are prone to brute-force dictionary attacks; biometrics systems require substantially more effort to successfully attack. Although standard encryption techniques are useful in many ways to prevent a breach of security, there are several new attack points possible in the domain of biometrics.

In remote unattended applications, such as web-based e-commerce applications, attackers may have enough time to make numerous attempts before being noticed, or may even be able to physically violate the remote client. At first glance, supervised biometric installations, such as at airports, may not be that vulnerable to brute-force attacks. But such installations surely can be the victim of replay attacks. Below, in Section 13.1 we develop a generic pattern recognition model that enables the study of security weak points. Such understanding is needed when designing biometric systems, while still keeping in mind the security *versus* convenience tradeoff.

A biometric authentication system looks like the system in Figure 40. An input device measures a biometric sample from a human being. This sample is converted into a machine representation of the sample (which can be the sample itself) by the "feature extractor;" this machine representation is matched against a reference representation, previously stored during enrollment. In Section 13.1, we describe the various attack points using a model like this.

13.1 Pattern recognition model

A generic biometric system can be cast in the framework of a pattern recognition system ([18, 134]). The stages of such a generic pattern recognition system are shown in Figure 40, indicated by A, B, C, D; enrollment is two stage: E and F. Excellent introductions to such automated biometric systems can be found in, e.g., [72, 106]. More recent descriptions can be found in reports like the *Biometric Device Protection Profile (DBPP)* [90, 158].

Any biometrics system is a four-stage system as in Figure 40. More specifically, these stages correspond to the typical stages:

A. The first stage is the actual acquisition of the input biometric sample. This process can be as simple as recording telephonic speech, to as intrusive as diluting the eye to acquire a picture of a prisoner's retina.

The actual process and logistics of this biometrics acquisition process is perhaps the most important factor in the convenience of a particular biometrics installation. On the other hand, the amount of control one has over the signal acquisition greatly influences the quality of the sample and therefore the accuracy of the system. Often the difficulty here is proper control of signal acquisition without overly inconveniencing the user.

B. The second stage is the processing of the biometric sample to obtain some digital machine representation that can be matched to other such representations. This stage can be as simple as just storing the biometrics sample, or it can be as complicated as computing complex index structures as in [57].

The problem of designing representations for biometrics is as difficult as designing machine representations for animate, articulated, and deformable objects, an active area of computer vision research

[53].

C. This stage is the computation of a probability of match between two or more representations, one representation of the input sample and one or more stored (enrolled) representations (from database F). There are several factors that determine how precisely this probability can be estimated for some input sample. These factors are the quality of the input, the precision with which the representation can be computed from a biometrics sample, and the effectiveness with which the matcher can compare two representations.

Another factor here is the quality of the training data that is used to optimize the matching process. Intricacies of this optimization process are discussed in Section 11.1

This stage also makes the actual *match* versus *no match* decision. This decision is best made with the particular application in mind.

D. This is the application, which is protected by the biometric authentication system. This can be a transportation system, a financial institution, or a computer account.

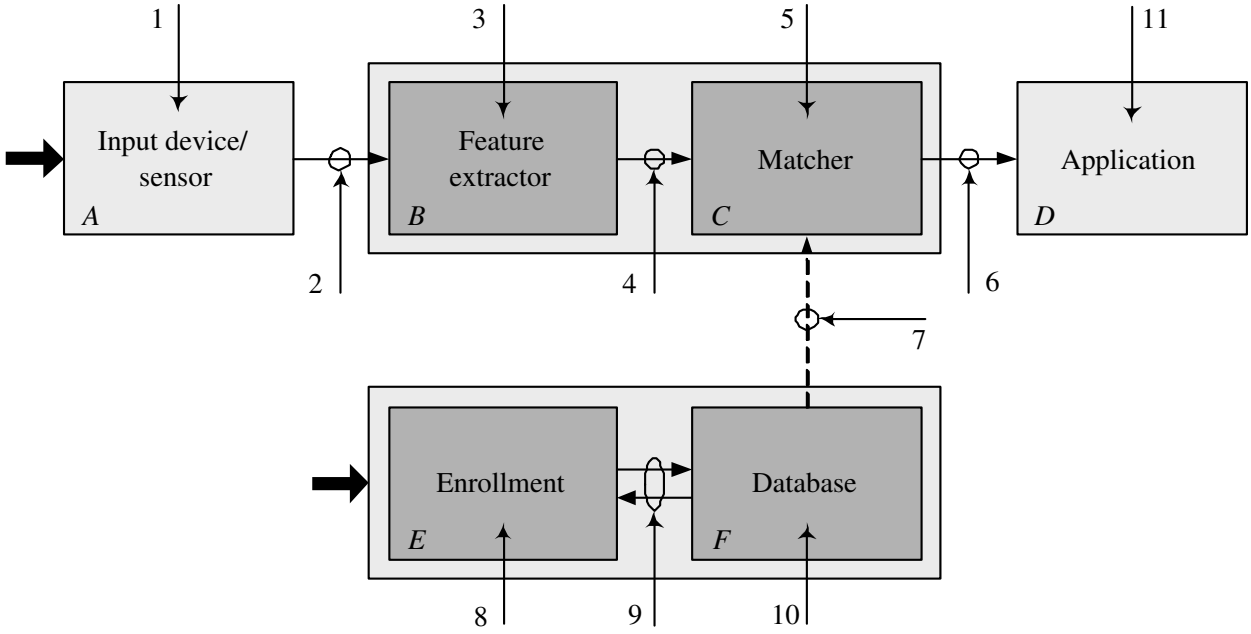


Figure 40: Stages of authentication system and enrollment system and points of attack in a generic biometric authentication system.

At this stage, the ultimate desired security of the decision, in terms of allowable False Accept (FA) rate, is chosen as the operating point, which gives the False Reject (FR) rate. This is a *compromise* between security and convenience, based on either static and or dynamic policies (see Section 3).

Besides the day-to-day work flow of a biometrics authentication or identification application, the implementation of the enrollment process and the organization of the database of biometrics samples are important aspects of the application. Specifically, referring to Figure 40, while enrollment E is more complicated than in password systems, a biometrics sample database F is similar to password storage.

E. The enrollment procedure and logistics is an often-underemphasized aspect of a biometric system. Unlike in password systems, it is not easy to re-enroll with a new biometric sample and therefore the enrollment process needs careful consideration. Section 12 discusses the intricacies of this part of a biometric authentication system.

F. The database of biometrics of samples F is either distributed, using smartcards [133] or centrally organized. This is a design choice in a biometrics system that can have many ramifications.

Within this model (Figure 40) we identify 11 basic points of attack that plague biometric authentication systems. We break these various attack points into several classes and expand on them below.

13.2 Violating biometric identifiers

Many attacks on biometric authentication applications are attacks based on somehow mimicking another person's authentication credentials. We list some of these attacks in this section. In terms of security, one could say that these are attacks on the integrity of the real-world actual biometric human appearance, like finger and face.

1. Threat 1 is the presentation of a fake biometrics at the sensor A, point of attack 1. This can be implemented as a replay attack by tampering with the sensor, or as an impersonation attack with some incidental amount of tampering with the sensory device or process (as described in Section 13.2).

This sort of violation can be achieved in two ways: mimicking a subject or physically attacking a subject. The most obvious case is when a genuine user is coerced to identify themselves to an authentication system. The authentication means could be forcibly extracted from a genuine user to gain access to the system with concomitant privileges. For instance, an ATM user could be forced to give away his or her ATM card and PIN at gunpoint. It is desirable to reliably detect instances of coercion without endangering the lives of genuine users and take an appropriate action when such coercion is detected.

The other possibility is that a biometric authentication application could be attacked by somehow changing one's biometrics. When cost is not an issue, all biometrics can be, and probably will be, subject to impersonation. The ease with which this can be achieved varies from very easy to close to impossible and *again* very much depends on the implementation of the specific system and, of course, the intrinsic error rates (Section 8) of the particular biometric.

For example, by using makeup or a mask it is easy to disguise one's facial appearance to make oneself look very different. This can prevent identification and so, in some sense, leads to false negative errors FN in biometrics identification applications (surveillance protocols). For instance, skilled humans have an uncanny ability to disguise their identity and are able to assume (forge/mimic) a different (specific) identity. To generate a FA error instead, one has to be able to both completely disguise one's true nature *and* faithfully impersonate (forge/mimic) a different individual. Face and voice are two traditional biometrics that come to mind first for this type of disguise or impersonation attack. However, every biometrics (except for perhaps DNA) is susceptible to such spoofing.

The security of a biometric system, when it comes to impersonation, depends on both the biometric and the particular installation. It will always be possible for one person to somehow impersonate others with a high degree of similarity (an important vulnerability in unattended, cooperative applications like physical access control). Photographs, rubber masks, and video replay all allow impostor attacks, i.e., allow active false accepts FA. Thus detection of security vulnerabilities using such fake biometrics data is important.

For example, there exist fingerprint scanners that automatically reject dead or fake fingers [146]. With computing power more abundant, the technology for detecting fake biometrics will likely keep improving. The combination of multiple biometrics also reduces the exposure to impersonation attacks since it means more attributes of an individual have to be copied with good fidelity.

As discussed in Section 6.1, the inclusion of biometrics in the authentication protocol, may inadvertently introduce points of attacks or security vulnerabilities. It are *these* points of attack that we study in the next section and, as we will see, biometric systems have many points of attack in common with password systems.

13.3 Front end attacks

The front end of the system is where the bulk of the authentication activity occurs. The front end is responsible for turning the sensed biometrics signal into some sort of invariant representation and then matching this against a retrieved reference template for the individual. This opens up several possibilities for attack:

2. A second point of attack of a biometric system is an attack on the channel between the sensor and the biometric system. We refer to this as threat 2. This can be a replay attack, resubmission of old digitally stored biometrics signal, or an electronic impersonation.

3. Threat 3 is attack by Trojan horse on the feature extractor B. The feature extractor could be attacked so that it will produce a pre-selected feature set at some given time or under some specific condition. That is, after the features have been extracted from the input signal they are replaced with a different synthesized feature set (assuming the representation is known).

4. Attack point 4 is the communication channel between the feature extractor and matcher. In the case of fingerprints, if minutiae are transmitted to a remote matcher (which can be the case when using smart cards [133] to store the template) then this threat is very real.

5. Attack 5 is again a Trojan horse, the matcher is attacked to directly produce an artificially high or low match score, thereby manipulating the match decisions.

13.4 Circumvention

An often overlooked vulnerability of an authentication system is:

6. Threat 6, a very important threat, is the overriding of the output of the matching module C. The output of the matching module could either be a hard *match* versus*no match* decision, or it could be just the probability of a match where the final decision is left up to the application. The attack point 6 is the same in both cases.

Some problems plague all authentication technologies (based on possessions, knowledge, or biometrics) alike. Fraud in an authentication system is possible in different forms. Some forms of fraud are characterized as loopholes in the system: possibilities of illegitimate access to a system not envisioned by its designers. Other forms involve using intentionally incorporated mechanisms to transcend the authentication used by the system (super-user) and hence, in principle, cannot be eliminated using any strategies embedded inside the system (intra-system). The types of fraud could be categorized as follows:

- *Collusion:* In any application, some operators of the system will have a super-operator status, which allows them to bypass the authentication component of the processing and to overrule the decision made by the system. This facility is incorporated in the system work flow to permit handling of exceptional situations, e.g., processing of individuals with no fingers in fingerprint authentication systems.

- *Covert acquisition:* It is possible that the means of identification could be compromised without the knowledge of a legitimate user and be subsequently abused. For instance, there is a significant amount of fraud in covertly observing PINs at public telephones.

  This could be called an *impersonation attack* as the ones discussed in Section 13.2, however, only a user's parametric data are used, the biometric is not impersonated. Yet it is unclear how one could realistically steal someone's fingerprint template by mere physical observation, nonetheless supply it to the authentication system. It is more likely that the attacker will duplicate some possesion or knowledge, such as a smartcard or PIN, which must be used in conjunction with the biometric (which has also been impersonated).

- *Denial:* It is possible that a genuine user may identify himself to the system using legitimate means of identification, through (say) a smart card, to gain access to the privileges and is subsequently denied such an access, i.e., a FR because of compromising biometric authentication templates. While this is not exactly fraud – no unauthorized access was granted to the protected resource – it disrupts the functioning of the system without explicitly breaking any of its components.

Many of these problems may not be fully eliminated. Currently, attempts to reduce fraud in authentication systems are process oriented and *ad hoc.* There is a need to focus research effort on systematic and technology intensive approaches to combat fraud in the system. This is especially true in terms of biometric authentication systems where the captured biometric measurements and context may have sufficient information to deter some forms of fraud. In particular, multiple biometrics show promise in approaching solutions to many of the above-mentioned problems.

13.5 Back end attacks

The database of enrolled individuals is available locally or remotely possibly distributed over several servers. Unauthorized modification of one or more machine representations in the database, which could result in authorization of a fraudulent individual, or at least in denial of service to the person associated with the corrupted template (again, it is assumed that the representation is known).

7. Threat 7 is another channel attack but on the communication between the central or distributed database and the authentication system. The (biometric) representations from the stored database F are sent to the matcher through a channel, which is attacked, to change the representations before they reach the matcher.

Processes E and F perform an extremely important function in a biometric authentication system, the enrollment of the eligible subjects, or the access control list. The "cleanness" of that database F is of extreme importance because the final authentication system is only as secure as its enrollment database itself (see Section 12). Three points of attack can be identified:

8. Attack point 8 is the enrollment center or application (E in Figure 40). The enrollment and authentication processes have similarities, and therefore enrollment is vulnerable at attack points $1, ..., 6$.

9. This point of attack is a channel 9 (similar to attack point 10). Control of that channel allows overriding the the (biometric) representation that is sent from F to C, assuming that the representation format is known to the intruder.

10. Attack on the the database F. The database of enrolled (biometric) representations is available locally or remotely possibly distributed over several servers. This threat is the unauthorized modification of one or more representations in the database. This could result in authorization of a fraudulent individual, denial of service to the person associated with the corrupted template (again, it is assumed that the representation format is known), or removal of a known terrorist from a screening list.

This also opens up the possibility of privacy attacks – an attack on the confidentiality of the biometric authentication system, i.e., on the access control list or database of members. This attack is not aimed at the application but is aimed at the biometric authentication system database. The section on privacy aspects of a biometric installation, Section 6.3, discusses already the possibilities and consequences of such attacks.

For example, with collusion between the hacker and the supervisor of the enrollment center, it is easy to enroll a newly created identity or stolen identity, the consequences of which could be severe. This threat is also very real in manual authentication systems. Purely from a biometrics point of view, the threat is ultimately related to the ease with which a biometrics can be impersonated (see Section 13.2) and hence to the intrinsic FA. Enrollment needs to be more secure than authentication and is best performed under trusted and competent supervision.

Of course, what also needs to be remembered is the application or system that is being protected:

11. As noted in [83], the actual application D is a point of attack, too. This means that biometric authentication systems should take advantage of all the security services that are offered in a traditional authentication system, too.

In addition, Schneier describes many other types of abuses of biometrics in [143].

Overall, the greatest threat to a biometric authentication systems is presenting, either physically or electronically, fake biometrics or previously acquired biometrics. This is especially a threat that needs to be addressed for enrollment, that is, careful thought must go into how easy it is to enroll a newly created identity. In particular, threat 1 and threat 2, somehow tampering with the input device or the communication channel so that the sensed biometric resembles the biometric of a target, need to be examined. As we note below, electronic impersonation is becoming, or will become, feasible. However, the threat of an electronic replay of previously recorded biometrics samples can be prevented using judicious combinations of challenge and response systems and data hiding (see [128] and Section 13.7).

The fact that biometric templates of the system in Figure 40 are stored either in a central database or distributed over clients or smart cards, does not necessarily affect the security of a biometric-based system since this portion of the system can be made secure with traditional technology. While a smart card solution may provide a privacy advantage since the biometrics database F is not centralized, they also give attackers ample time to tamper with the smart card.

The issue of privacy raises many concerns and needs to be solved by designing identity-encrypting technology. In that fashion, identity encrypted templates can be matched in the "encrypted" domain without

any possibility of ever tracing back the original identities. (A first attempt to this can be found in [128].) The detection of fake biometrics is also a very important research topic that has not been addressed sufficiently yet.

13.6 Other attacks

Password systems are vulnerable to brute force attacks, here the number of characters in the password corresponds to the bit-strength of the password, which expresses the amount of effort it takes (on average) to break into password authentication systems. There exists an "equivalent" notion of bit-strength of a biometric, also called *intrinsic error rate* (see Section 8). Such attacks would occur either at Point 2 or Point 4 but, in general, the number of variations that would have to be tried is prohibitive.

In Section 8 we discuss some of the motivations for research in the bit strength of biometric identifiers. More on this subject can be found in [131, 134].

Attacks that we have not mentioned yet but are mentioned in [90] are:

- "*Hill climbing attack:* A type of malicious attack directed towards the comparison process whereby an unauthorized user incrementally increases the proposed matching data and presents this data directly into the comparison function until a successful matching score is provided by the biometric algorithm." This would be similar to a brute force attack. For instance, for fingerprints the attacker might submit a print with hundreds of minutiae in the hopes that at least the threshold number N of them will match the stored template (assuming the matcher does not reject such represnetations or normalize for this effect).
- "*Piggy-back attack:* A type of malicious attack whereby an unauthorized user gains access to the protected assets through simultaneous entry with a legitimate user. This attack may be characterized by physical force or logical entry beyond the portal." This corresponds in some ways with coercion as mentioned in attack Point 1.

13.7 Challenge and response

One of the prime vulnerabilities of biometric authentication systems is replay attacks, particularly at Point 2 and Point 4. One way to guard against these is to use a challenge-response protocoal. For example, prompted-text or text-independent verification (which are well established in speaker identification literature [4]) can avoid a simple replay attack, but at the cost of a more intrusive, complex and expensive system. Ironically, advances in trainable speech algorithms [43, 58] provide tools for attacking even these more sophisticated systems. Such challenge and response systems can be extended to other biometrics. Interactive fingerprint authentication systems can use video streams of fingerprints [44]. Of course, even these systems will eventually be attacked successfully. It is difficult to predict whether man will beat machine, or *vice versa*. However, by making impersonation more difficult, at least casual attacks (the majority) can be deterred.

Another type of challenge and response to detect replay attacks in hardware is proposed in [131]. This approach is based on challenges to the sensor that is assumed to have enough intelligence to respond to the challenges. Many silicon fingerprint scanners [7, 140] are able to exploit the proposed method as a local secure processor can be integrated without much effort.

These type of challenge and response implementations, either between human and computer, and computer and computer are naturally part of the authentication protocol $R(P, K, B)$ and fall in the category of a dynamic authentication protocol as described in Section 14.3.

14 Integrating information

For many applications there are additional sources of information that can be used in authentication. In other applications, the use of a single biometric is not sufficiently secure. In both cases the question becomes how does one integrate multiple sources of information to make the whole application more secure. Below we describe a variety of techniques for doing this based on different sorts of information.

14.1 Searching biometric databases

Some of this hype is surrounding the accuracy of (large-scale) biometric *identification*, i.e., $1:n$ biometric matching or searching a biometric database $\mathbf{N}$ with n enrolled subjects. The optimal way to implement such a search is by performing n biometric $1:1$ matches, either sequential or in parallel. However, these searches are often constrained to search smaller portions of the database.

For databases $\mathbf{N}$ holding a large number (n) of biometric templates, "search efficiencies can be achieved by partitioning them into smaller groups based both upon information contained within (endogenous to) the templates themselves and upon additional (exogenous) information such as the customer's name, obtained at the time of enrollment" [169]. These techniques are called "filtering" and "binning," constraining the search with parametric data and constraining the search with additional biometric data.

Filtering: Examples here of course are plenty. Filtering down the search of $\mathbf{N}$ by (say) a subject's surname, could partition the biometric template database into many sets, which can be overlapping when fuzzy comparison between names matches is used (some sets can be very large, c.f., "J. Smith"). Note that filtering then is an authentication protocol P that operates on authentication method, or credentials, $(possession, biometric) = (P, B) = (name, B)$. The protocol simply says:

1. Search the database $\mathbf{N}$ of enrolled subjects by surname (possibly a fuzzy search).
2. Search the subjects with matching surnames by matching input sample biometric B with the corresponding templates.

This returns a "candidate list" or "hit list" of subjects in database $\mathbf{N}$ with the name *Smith* and a template that matches biometric sample B.

The extreme here is narrowing down the search using a unique (or almost unique) parametric identifier, i.e., social security number or some other national identifier. Then the credentials are

$$(possession, biometric) = (P, B) = (sosial\ security\ number, B)$$

and the problem of identification (1 : n matching) is reduced to verification (1 : 1 matching).

Binning: This is filtering based upon information contained within the templates in database $\mathbf{N}$ themselves, or more general, based upon other biometric identifiers B' of an individual. Such a search is performed by an authentication protocol operating of an authentication method $(B) = (\{B, B'\})$. Perhaps the best known instance of this technique is to *first* classify the type of the fingerprint and *then* matching the minutiae of the fingerprint. Here the types are global print classes like "loop," "whorl," etc. and the minutiae are much finer resolution features [51].

Binning can also be based on different biometrics of an individual, i.e., $(B) = (\{B, B'\})$. Here B could be a fingerprint and B' could face type, which may be obtained from a face image. If the search space (the number of matches) is to be reduced with this authentication method is the protocol:

1. Select those subjects in the database **N** whose biometric template matches biometric (sample) B', i.e., those subjects with face type B'.
2. Match the input credential B with the templates of these remaining subjects to find those subjects in with matching B' and B.

Which ideally returns one unique enrolled subject from database **B** or a small list.

This type of binning is commonly referred to as "multiple biometrics" and creates the problem of biometric matcher integration as discussed in Section 14.2

When using the filtering or binning authentication methods (P, B) or $(\{B, B'\})$, the space to which a subject biometric $\mathcal{B}$ is matched is limited to a smaller part of the database of templates, i.e., the protocol is directed at limiting the number of biometric $1:1$ matches. In a sense, the database is divided up into multiple partitions; these partitions may overlap in case of uncertainties in templates; a template may be in all partitions. Then the percentage of the database that is scanned on average P_e is called the penetration rate

$$\mathrm{P}_e = E(\#\, of\ biometric\ matches)/n.$$

Filtering is another type of integration, that of authentication protocols based on integrating possession type P of credentials, such as $name$, $home\ state$, etc. to be used to assist and speedup the biometric search. Possession tokens can be added to an existing authentication protocol:

1. Adding P to the *identification protocol* prescribes narrowing down on a smaller and smaller set of biometric templates. This has the desirable effect that the chances of false positives go down, the probability of a false negative, will go up dramatically. A matching biometric template in the database **N** that is not associated with the additional possession identifiers will *always* produce a false negative.

2. Adding P to a *verification protocol*, asking for additional possession beyond the identifier ID. Note that this type of integration does not decrease the *biometric* verification error rates and, therefore, does not the increase the probability that the subject is who he claims to be. After all, the probability that a person is who he says he is equals the probability of match between a stored biometric template and a newly acquired biometric sample. This probability does not increase when the subject supplies additional passwords, it simply means that the subject knows the passwords.

   To form a true composite probability, one would have to estimate $p(Pw/\neg s)$, the probability that someone other than subject s knows password Pw.

14.2 Integration of biometrics

One of the primary limitations of biometrics are the often higher than desired error rates and the use of multiple biometrics is thought to be a solution to existing authentication protocols.

1. *In an identification protocol* an additional biometric might be used to narrow down on a smaller and smaller set of biometric templates. Hence, the authentication protocol calls for first matching B and then matching B', hence only if both B AND B' match the template, a subject is considered a match for the candidate list.

2. *The verification protocol*

The use of more than one biometric may alleviate problems, or adverse properties, of the individual biometrics as described in Section 10. These are properties of a biometric like universality, collectability, and acceptability; for example, an additional biometric beyond finger accommodates those subjects without fingers or with hard to acquire fingerprints and accommodates those who object to enrolling with a fingerprint.

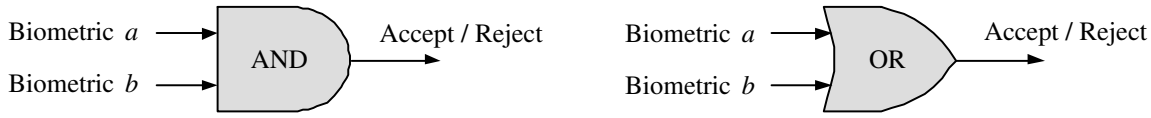


Figure 41: Verifying that *both biometrics* match or *one of the biometrics* matches.

We formulate the problem of multiple biometrics integration and examine some integration schemes. The actual mechanics of integrating the authentication using multiple biometrics (say, biometric a and biometric b) is specified in the authentication protocol. Figure 41 shows simple biometric integration at the decision level. For instance, let biometric a be a finger of a subject and biometric b be the face appearance of the subject. Also let the false accept and false reject rate for biometric a be FA_a and, FR_a; and for biometric b, FA_b and FR_b.

The effect that the integration of multiple biometrics can have on the overall FA_o and FR_o of the integrated system. Combining the biometrics through an AND improves the FA at the expense of FR; combining the biometrics through an OR improves the FR at the expense of FA. Again, this is a tradeoff between security and convenience. This can be seen as follows:

- *Biometric a* AND *biometric b:* To generate a false accept, both biometrics a and b have to be falsely accepted, hence
$$\mathrm{FA}_o = \mathrm{FA}_a\mathrm{FA}_b;$$
to generate a false reject either biometric a or b needs to be falsely rejected, hence
$$\mathrm{FR}_o = \mathrm{FR}_a + \mathrm{FR}_b(1 - \mathrm{FR}_a) = \mathrm{FR}_a + \mathrm{FR}_b - \mathrm{FR}_a\mathrm{FR}_B.$$
Here the product is subtracted to avoid counting false rejects FR_b when FR_a is true.

- *Biometric a* OR *biometric b:* To generate a false reject, both biometrics a and b have to be falsely rejected, hence
$$\mathrm{FR}_o = \mathrm{FR}_a\mathrm{FR}_b;$$
to generate a false accept either biometric a or b needs to be falsely accepted, hence
$$\mathrm{FA}_o = \mathrm{FA}_a + \mathrm{FA}_b(1 - \mathrm{FA}_a) = \mathrm{FA}_a + \mathrm{FA}_b - \mathrm{FA}_a\mathrm{FA}_b.$$
Table 20 summarizes the error rates, assuming that the product $\mathrm{FA}_a\mathrm{FA}_b$ is small compared to the sum of the error rates.

The effect of integration of biometrics on the error rates is better understood if concepts like the Receiver Operating Characteristic (ROC) curve are defined. Therefore, further discussion of this topic is deferred to Section 14.4. An example of a system that uses two biometrics can be found in [64].

	FA_o	FR_o
AND	$\mathrm{FA}_a\mathrm{FA}_b$	$\mathrm{FR}_a + \mathrm{FR}_b$
OR	$\mathrm{FA}_a + \mathrm{FA}_b$	$\mathrm{FR}_a\mathrm{FR}_b$

Table 20: Biometrics can be integrated to improve security (AND) or to improve convenience (OR).

14.3 Dynamic authentication protocols

Protocols do not have to be static, indeed dynamic protocols have been introduced. One dynamic protocol for speaker verification is the idea of *conversational biometrics*. Quoting from [16]:

> "... A conversational biometric systems engages in a natural language dialog with the user, presents randomly generated questions, and collects the users responses. When appropriate, the conversation may be embedded into the natural transaction dialog. The questions to be asked can be randomly generated from a large collection of questions and answers that the user provides during an explicit knowledge enrollment stage (e.g., "what is your favorite color?"), or may be generated from user-specific information available from the application (e.g., "what is the balance in your last statement?"). A protocol management module generates a protocol dynamically for each transaction, based on the application requirements. The dynamic protocol specifies the maximum number of questions to be asked, and the minimum number of correct answers, or alternatively, the protocol may specify the minimum score needed for the knowledge match. The protocol also specifies the minimum score required for the acoustic voiceprint match, and may adaptively modify the maximum number of questions to be asked, based on the voiceprint match scores. By generating appropriate protocols, one can provide higher accuracy or greater flexibility, or both [98, 113]."

Note that conversational biometric does a biometric match between the speech sample and the voiceprint on file and a "knowledge match" between the collected responses through speech recognition and the knowledge on file. Verifying the correctness of the answers however does not increase the probabilistic certainty about a subject at the other end of the phone line; it may increase *the possibility* that the subject is authentic but it does not increase the probability that the subject is authentic. This is because the answers are not inherently associated with only the subject and *not* provably unknown to others, i.e., known to possible impersonators and imposters. Hence, such authentication has repudiation possibilities as a liability.

Referring back to Section 3 where the three methods of authentication, *possession, knowledge,* and *biometrics*, are defined. The answers to questions like: *What is your mother's maiden name?* and *How many miles do you have accrued?* are knowledge tokens K. We can again view conversational biometrics as an authentication protocol

$$R(P, K, B),$$

with R operating on the triplet (P, K, B).

Let us look at a list of questions and the answers for the conversational biometric of Table 21. These questions have to be somehow agreed upon during an enrollment phase, there are two possibilities here: (i) questions/answers about private or personal issues but about issues that are not really secret; (ii) questions/answers about issues that are somehow secret. In both cases, the conversational answers are knowledge tokens K. Conversational knowledge may appear to be a possession P (the accrued mileage) and is often confused with biometric information B (mother's maiden name) because it has many of the properties of

Conversational biometrics	
Place of birth	Virginia
Name of pet	Tiger
.	.
.	.
Q_k	A_k

Table 21: A list of questions and answers that represents the knowledge K in conversational biometric authentication.

biometrics. Conversational biometric knowledge K are like passwords and phrases however because the knowledge K can be shared. Either way, the authentication protocol $R(P, K, B)$ calls for simultaneously recognizing a subject based on voice B while verifying the answers to the questions K.

Conversational biometrics could use other forms of man-machine communication, for example using handwriting recognition or keystroke characteristics (which is a behavioral biometric). This could be signature verification generalized to writer identification [65] with the same categories (as for speaker verification), i.e., *text dependent* and *text independent* writer verification. As a working biometric technology, attention has focussed on signature verification and writing verification as areas of research. Authentication protocols where the subjects are asked to provide written or typed answers to questions can be envisioned; there is evidence here that a person's signature and person's handwriting are complementary [151]. Of course, one can always use a question bank as in Figure 21 with any biometric.

In effect, conversational biometrics is an extension of machine to machine secure authentication protocols, called challenge and response protocols, see Section 13.7. Here, electronic inquiries that can only be answered by an authentic system are dynamically formulated.

Many traditional biometric authentication protocols are dynamic in nature. Interactive input sample quality control implemented through communication between human and computer is for instance a dynamic protocol, see Section 12.4.

14.4 Information integration

The above parametric properties, $\mathrm{FA}, \mathrm{FR}, \mathrm{FTE}, ...$ are studied for biometric authentication protocols that involve one biometric identifier. The integration of biometric identifiers is an *integral part* of the authentication protocol. The design of multiple biometrics authentication systems with specified FA and FR is quite complicated as we see in the rest of this section.

14.4.1 Integration methods

Information contained in multiple biometrics could be integrated using a number of different methods, at various levels, and in different contexts (see, for instance, Figure 42).

The various types of integration indicated in Figure 42 are described in the following and example protocols are given in Table 22.

1. Multiple biometrics: This is what typically is thought of as integrated biometrics authentication. More than one, disperse biometrics, such as face and voice, or fingerprint and hand geometry are used for

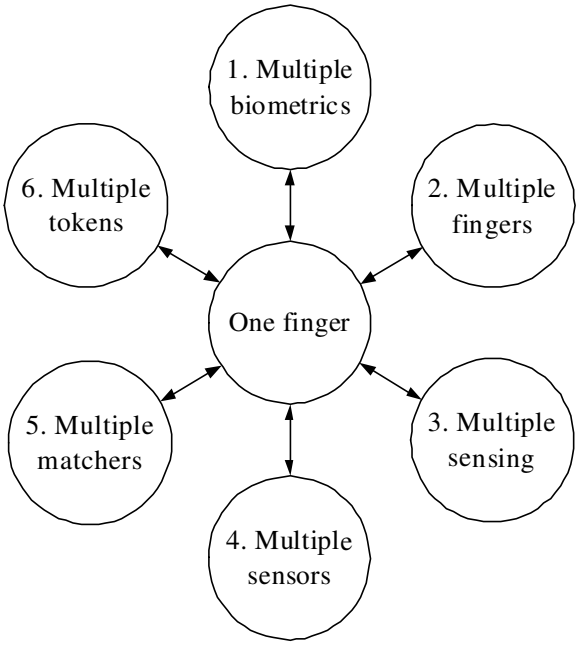


Figure 42: Multiple choices for integration in a fingerprint system.

authentication. This can be through the AND-protocol or the OR-protocol.

2. Multiple locations: This authentication protocol is only possible with biometric identifiers that are present on the subject in multitude. Examples are the left OR the right iris, the left index finger AND the right index finger, and so on.

3. Multiple sensing: This includes, for example, the well-known authentication protocol "three tries and you're out." That is, a subject is allowed to offer three samples of the biometric and has three chances to match this sample with the stored templates. The variants here are for instance giving the subject some upper time limit for voice and face recognition.

4. Multiple sensors: Two or more different sensors can be used to acquire the same biometric identifier, e.g., a fingerprint sensor based on frustrated total internal reflection and a sensor based on CMOS technologies. Such protocols include the use of multiple cameras from different viewpoints.

5. Multiple matchers: Here different matching technologies are used on the same biometric sample acquired from the subject. These are the combinations of feature versus appearance representations (templates) for face images and fingerprint image, where the features would be minutiae.

6. Multiple tokens: For completeness we have added the integration of a biometric with a possession or knowledge token. Again, this is the typical authentication with biometric verification protocol as discussed in Section 3.1.

14.4.2 Score level integration

There are many ways to integrate biometric identifiers:

	Example authentication protocol
1. Multiple biometrics	Face image and voice print.
2. Multiple locations	Index and middle finger.
3. Multiple sensing	Three tries of index finger.
4. Multiple sensors	Ultrasonic and optical sensing.
5. Multiple matchers	Minutiae and correlation fingerprint matcher.
6. Multiple tokens	Tokens as in Section 3.1

Table 22: Fingerprints lend themselves to many different types of integration.

Tightly coupled integration: The output from multiple biometric sensors could be used to create a more reliable and/or extensive (spatially, temporally, or both) input acquisition [24]. The representations extracted from many biometric sensors could be collated and the decisions could be made based on the joint feature vector. The integration at sensor or representation level assumes a strong interaction among the input measurements and such integration schemes [32].

Loosely coupled integration: On the other hand, assumes very little or no interaction among the inputs (e.g., face and finger) and integration occurs at the output of relatively autonomous matching engines, each engine independently assessing the acquired biometric samples.

This document only discusses loosely coupled integration systems.

The loosely coupled systems are not only simpler to implement, they are more feasible in commonly confronted integration scenarios. A typical scenario for integration is two biometric systems (often proprietary) independently acquiring inputs and making an autonomous assessment of the "match" based on their respective biometrics; while the decisions or scores of individual biometric systems are available for integration, the features used by one biometric system are not accessible to the other biometric system.

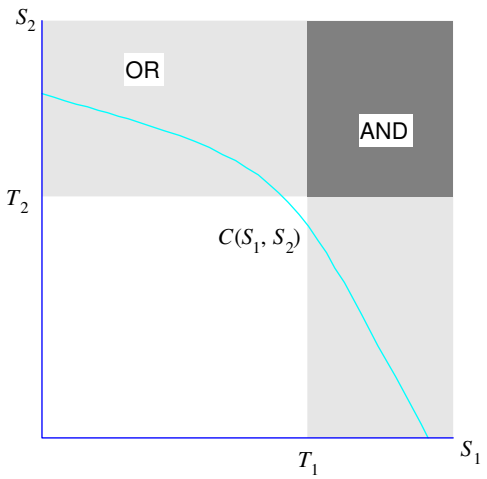


Figure 43: One decision region is admissible under the OR-rule; another region, contained in the first region, is admissible under the AND-rule.

Decision level integration: Here, as in Figure 41, only decisions made by the individual biometric matchers are integrated. Let the match scores for Biometric 1 and Biometric 2 be s_1 and s_2 for a particular subject, respectively.

The AND *integration protocol:* A subject is required to offer *both* the biometric identifier 1 and identifier 2 to the biometric authentication system. The subject is authenticated if $(s_1 \geq t_1)\ AND\ (s_2 \geq t_2)$. In Figure 43 this means that a subject is authenticated if the score pair (s_1, s_2) lies in region **I**.

The OR *integration protocol:* The subject is asked to offer the *first* biometric identifier 1 and *then* identifier 2. A subject is authenticated if $(s_1 \geq t_1)\ OR\ (s_2 \geq t_2)$ where the second biometric only needs to be acquired if $s_1 < t_1$. In Figure 43, a subject is authenticated if the pair of scores, (s_1, s_2), lies in region **II**.

The OR protocol can be implemented in various ways, in terms of the order of acquiring biometrics 1 and 2.

Score level integration: Here a subject is authenticated if $C(s_1, s_2) \geq t$, i.e., if the score pair (s_1, s_2) lies above the curve $c(s_1, s_2) = t$ as in Figure 43.

Optimal integration then amounts to finding that curve $c(s_1, s_2)$ at which (say) the FA rate is below some desired minimum level and the FR is as low as possible. Hence, the authentication system is operating at a curve, rather than at a point on the ROC (see quantitative parameters listed in Section 5). The curve could be found by using test databases of multiple biometric samples. The literature does not appear to reveal any research on this important topic. Below, we present a glimpse of the complexity in determination of a desirable operating curve. The choice of operating point for single-biometric systems is discussed in Section 5.1.

Note that this optimization may include the degenerate solutions. For example, if we look at Biometrics 1, it may be found that $s_1 = 0$ or $s_1 = 1$, which means that Biometrics 1 is irrelevant in the final authentication decision. More specifically we have:

$(0, c(0, s_2)) = (0, t)$ and $(1, c(1, s_2)) = (1, t)$. Operating at these degenerate thresholds ($t = 0$ and $t = 1$) means:

1. $s_1 = 0$: We have authentication protocol, $c(s_1, s_2) \geq t$, and we get $c(0, s_2) = s_2 > t$. This corresponds to the *AND* rule where Biometric a is ignored by setting its matching threshold to zero and thereby accepting every subject irrespective of match score of Matcher 1.

   The "integrated" system operates just on the match score s_2 from Biometric 2.

2. $s_1 = 1$: Again, with authentication protocol, $c(s_1, s_2) \geq t$, we get $c(1, s_2) = s_2 > t$. This corresponds to the *OR* rule where Biometric a is ignored by setting its matching threshold to one, which means that nobody is accepted based on the score of Matcher 1.

   The "integrated" system again operates just on the match score s_2 from Biometric 2.

Hence, it is *not necessarily true* that integrating multiple biometrics gives better error rates. There is no research evidence available as to whether these degenerate optimal solutions actually can occur when trying to integrate the common (see Section 4) biometrics *or not.*

15 APIs, standards and databases

A rapidly developing technology like biometrics needs to develop standards to address issues of interoperability, plug-and-play compatibility and building common Application Program Interfaces (APIs). In a typical biometrics recognition system there are three layers of interaction involved as shown in Figure 44. The lowest layer is the hardware layer involved with interfacing with the biometrics-specific hardware. Also at the hardware level, standards like what ports and connectors the devices should use are being developed. The next level deals with the recognition of the basic biometrics signal with vendor-specific representation templates. The highest level is the layer involving the application. Standards are needed at every layer. In this section we briefly present the existing standards for these layers.

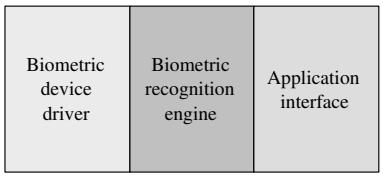


Figure 44: Layers of interaction with biometric authentication systems.

For wide acceptance of biometrics, standards for interfaces and performance evaluation are needed. Several standards are in the process of being developed and promoted. NIST is playing an important role in designing several fingerprint databases [163, 164, 165, 166, 167] and conducting speaker verification tests. The US Department of Defense runs the FERET and FRVT face recognition test. The BioAPI [8, 156] is a standard for the application programmer interface allowing the decoupling of biometrics-technologies from the applications that use them. At the hardware level, the devices for biometrics still remain non-interoperable except when sharing a common existing standard such as NTSC video.

15.1 Standards and recommendations

We mention some of the better known standards and the standards known to the authors. We do not guarantee that this list of standards is exhaustive.

- *ANSI-NIST-ITL-2000:* This standard specifies a common format to be used to exchange fingerprint, facial, scars, mark and tattoo identification data effectively across dissimilar systems made by different manufacturers. This standard [54] is a revised version of the ANSI-NIST-CSL-1993 standard and ANSI-NIST-ITL 1A-1997.

- The *1996 FBI WSQ* standard: For compression and decompression of gray scale fingerprint images, the FBI in collaboration with LANL has developed a wavelet-based standard known as Wavelet Scalar Quantization (WSQ). The original standard was proposed in 1993 and revised in 1996. The significant advantage of the standard is that even at higher compression ratio, the minutiae feature are less disturbed compared to other image compression standards. The law enforcement agencies where storage of the fingerprint images is required, WSQ plays a crucial role.

  More details about WSQ are available in [52].

- *AAMVA DL / ID 2000:* The American Association for Motor Vehicle Administration (AAMVA) Drivers License and Identification (DL/ID) Standard [1] describes a uniform means to identify issuers

and holders of driver license cards. The standard specifies identification information on driver license and ID card applications. It also specifies a format for fingerprint minutiae data that would be readable across state and province boundaries for driver licenses.

- *CBEF:F* The common Biometrics Exchange File Format (CBEFF)is an emerging standard that deals with the issue of sharing biometrics in raw signal form or in template form. The standard is recommended for facilitating data interchange between different components of a system or even across systems. The data described by CBEFF includes security information in the form of digital signatures and data encryption; identification of biometric type and information about the biometric sample and the actual biometric data.

  More details about CBEFF are available in [28].

- *ANSI X9.84:* For use of biometrics in the financial industry, X9.84 describes a standard for biometric information management and security. The X9.84 standard defines requirements for managing and securing biometric information such as customer identification and employee verification. X9.84-2000 specifies the minimum security requirements for effective management of biometrics data for the financial services industry and the security for the collection, distribution and processing of biometrics data. The standard includes specification for the security of the physical hardware used, the management of the biometric data, the utilization of biometric technology for verification/identification of banking customers and employees, the application of biometric technology for physical and logical access controls, the encapsulation of biometric data and techniques for securely transmitting and storing biometric data. The biometric data object specified in X9.84 is compatible with CBEFF.

  More details about X9.84 are available in [2]

- *Intel's CDSA:* Intel's Common Data Security Architecture (CDSA) standard provides a set of security building blocks for application developers. CDSA is designed as an overall infrastructure for data security on PCs, workstations, and servers. It is founded on two fundamental data security premises: digital certificates (a form of electronic identification that enables a hierarchy of trust, dependent on the identity of the user) and portable digital tokens, which store cryptographic keys and perform cryptographic operations. The CDSA 2.0 specification introduces Elective Module Managers (EMM) as a key CDSA component. An EMM can add new and compelling security features such as biometric authentication.

  Human Recognition Services (HRS) is an EMM under CDSA. Its basic purpose is to verify the identity of a person based on some combination of password knowledge and biometric measurement. The HRS EMM supports the integrity features of the CDSA stack which are currently not provided by the BioAPI. More details about CDSA can be found in [66].

15.2 Application program interface

There is much work ongoing in the area of standardization of Biometric Application Program Interfaces (APIs), at all levels of interaction of Figure 44. Given all the possible biometric protocols, in particular the dynamic conversational protocols, obviously this is quite a complicated task. We mention a few API standardization efforts:

- *BioAPI:* The BioAPI is the most popular API in the biometrics area. It is intended to provide a common interface for authentication based on any biometrics. The primitives included in BioAPI

allow applications to manage enrollment, verification, and identification tasks on a client/server and the biometrics signal acquisition on a client.

At the highest level, BioAPI defines a Biometrics Service Provider (BSP) which deals with all the aspects of biometrics signal processing. The three steps in the processing are: capture, process, and match. Any biometric data that is returned to the application is referred to as Biometric Identification Record (BIR). The three main classes of functions supported by the API deal with enrollment, verification, and identification. BioAPI components post information about themselves in the BioAPI module registry during installation. The module registry can be used by applications to check the BSPs installed and their functionalities. Any device specific parameters including the status of the device can also be contained in the registry to let the application decide the correct device and its usage.

More details and reference implementation code are available in [8].

- *BAPI:* This is an API advocated by I/O Software Inc. [67]. This is an API that allows the programmer to develop applications for a broad range of Virtual Biometric Devices (VBDs) without knowing the specific capabilities of the device. The API is comprised of three distinct levels of functionality from high device abstraction to low (device specific) abstraction.

  BAPI is a standard s/w protocol and API for communication between software applications, the operating system and biometrics device. Individual hardware and related modules can easily be swapped in or out. This API is device agnostic. There is no consortium for BAPI. Any software/hardware developer, integrator or solutions provider wishing to integrate with BAPI and Microsoft Windows should contact I/O Software [10].

15.3 Databases

For impartial performance evaluation of biometric recognition systems, the industry needs to have access to large public databases. In this area, NIST has been playing a leader for a long time with a large variety of databases for fingerprints, voice samples, and mug shots. With significant interest in the face recognition, there are now several large databases, many collected by academic institutions. For speaker recognition, there are also additional academic databases available publicly. However, for iris *no known public database* is available at this time.

15.3.1 Face databases

Because of forensic applications of face recognition, interest in face databases has been in the mission of NIST all along; more recently, because of the tremendous interest in face recognition in academia, an abundance of university face databases is emerging:

- *NIST 18 Mug shot Identification Database* (http://www.nist.gov/srd/nistsd18.htm): The NIST 18 database consists of 3248 mug shot images for testing mug shot recognition systems. The database contains both front and side (profile) views when available. Separating front views and profiles, there are 131 cases with two or more front views and 1418 with only one front view. The images have been scanned at a resolution of 500 dpi.

- *FERET* (http://www.dodcounterdrug.com/facialrecognition): The Facial Recognition Technology (FERET) database was collected over several sessions spanning over three years. The database contains 1564

sets of images for a total of 14,126 images that includes 1199 individuals and 365 duplicate sets of images. A duplicate set is a second set of images of a person already in the database and was usually taken on a different day.

For some individuals, over two years has elapsed between their first and last sittings, with some subjects being photographed multiple times. More exhaustive face databases are now being made available through FRVT 2002. More details about how to get a copy of FERET database are available at [119].

- *FRVT 2002* (http://www.dodcounterdrug.com/facialrecognition): The main idea of the Facial Recognition Vendor Test (FRVT) 2000 was to evaluate face recognition technology performance in the real world. The FRVT 2000 used most of the FERET database. The FRVT 2002 extends the test to cover more difficult databases including a short video of faces. The largest face database to be used in the test involves 120,000 faces. More details about the data will be made available at http://www.frvt.org/FRVT2002/default.htm.

- *M2VTS* (http://www.de.infowin.org/ACTS/RUS/PROJECTS/ac102.htm): Other than the government initiated programs for face databases, M2VTS (M2VTS Project: Multi-modal Biometric Person Authentication) has been an active player in face databases. In fact it is the only multi-modal database that includes face and speech. The database consists of 37 persons and faces taken at 5 different orientations. Later this database has been extended to a large face video database and called *XM2VTSDB*. The new database consists of 295 persons and four sessions taken over several months. High quality color images are available.

- *The CMU database* (http://www.ri.cmu.edu/projects/project_418.html): There are several different face databases available from the CMU Robotics Institute. The CMU Pose, Illumination, and Expression (PIE) database consists of 41,368 images of 68 people. Each person face has been imaged under 13 different poses, 43 different illumination conditions, and with 4 different expressions. The other databases include facial expression analysis database and face detection database. More details about these databases are available at the web site mentioned above.

- *The Yale database* (http://cvc.yale.edu/projects/yalefacesB/yalefacesB.html): The Yale Face database B contains of 5850 single light source images of 10 subjects each imaged under 576 viewing conditions arising from 9 poses and 64 illumination conditions. For every subject in a particular pose, an image with ambient (background) illumination is also captured. The acquired images are 640(w) $\times$ 480 (h) in size and have 256 gray scales. Even though a large number of poses are available, the database has only 10 subjects. Hence, it is better suited for face modeling.

- *The MIT database* (ftp://whitechapel.media.mit.edu/pub/images/): The MIT face database is one of the oldest public face databases. It consists of faces of 16 subjects and 27 images per subject with varying illumination, scale, and head orientation. The data set can be directly downloaded from the above site. For large scale system testing, this database may not be suitable.

- *The Purdue database* (http://rvl1.ecn.purdue.edu/ aleix/aleix_face_DB.html): A relatively large face database is available from Purdue University. It contains over 4,000 color images corresponding to 126 subjects imaged with different facial expressions, illumination conditions, and occlusion. Each person participated in two sessions, separated by two weeks time. The images are 768×576 pixels with 24 bit color. The database is available freely for research purposes from the URL specified above.

15.3.2 Fingerprint databases

As mentioned NIST has been very active in the area of fingerprint database collection. These are databases with prints from digitized paper impressions, annotated fingerprint classes, and fingerprint image sequences. Information about the NIST databases is found at http://www.nist.gov/srd/ but we give a brief description here:

- *NIST-4* (http://www.nist.gov/srd/nistsd4.htm): The NIST-4 fingerprint database is a very popular database consisting of 2000 fingerprint pairs scanned at 500 dpi. Each image is 512×512 pixels and has a class label. The database consists of 400 image pairs from each of the 5 classes. Even though this database is more suitable for classification tests, researchers have been using it for testing matching performance also.

- *NIST-9* (http://www.nist.gov/srd/nistsd9.htm): The NIST-9 is one of the largest rolled fingerprint databases consisting of 5 volumes. Each volume has 270 mated fingerprint cards resulting in 5400 images. The cards have been selected to reflect the natural fingerprint distribution. Each image is 832 by 768 pixels scanned at 500 dpi resolution and the National Crime Information Center (NCIC) class marked by FBI is available.

- *NIST-10* (http://www.nist.gov/srd/nistsd10.htm): The NIST-10 database is a supplemental database to NIST-9 consisting of a larger sample of fingerprint patterns that have a low natural frequency of occurrence and transitional fingerprint classes. The database consists of 5520 images. Each segmented image is 832 by 768 pixels scanned at 500 dpi resolution. The NCIC class provided by FBI is also available. This database is more suitable for algorithm testing involving rolled fingerprints.

- *NIST-14* (http://www.nist.gov/srd/nistsd14.htm): NIST Special Database 14 is suitable for development and testing of automated fingerprint classification and matching systems on a set of images which approximate a natural distribution of the fingerprint classes. Each segmented image is 832 by 768 pixels, scanned at 500 dpi and classified using the NCIC classes given by the FBI and WSQ compressed. The full database consists of 27,000 fingerprint pairs.

- *NIST-24* (http://www.nist.gov/srd/nistsd24.htm): There are two components in NIST-24 livescan fingerprint video database. The first component consists of MPEG-2 compressed video of fingerprint images acquired over 10 seconds when the user intentionally distorts the impression. Each frame is 720×480 pixels. All ten fingers of 10 subjects are available. This database is useful for studying impact of distortion on matcher performance as well as distortion detection. The second component of the database deals with image acquisition at various rotations. Similar to the first component, it consists sequences of all ten fingers of 10 subjects. The MPEG-2 compression for both the databases have been carried out at 5 Mbits/second.

- *NIST-27* (http://www.nist.gov/srd/nistsd27.htm): In the area of latent scene-of-crime fingerprint analysis, NIST-27 is a very useful database. It consists of 258 latent fingerprints from crime scenes and their matching rolled fingerprint mates with the minutiae marked on both the latent and the rolled fingerprints. All the paired minutiae between the latent and its corresponding rolled print are also identified. This database is useful for testing latent matching performance.

- *NIST-29* (http://www.nist.gov/srd/nistsd29.htm): The NIST-29 fingerprint database consists of both rolled and plain fingerprints of the same fingers that can be used for system performance testing. The

images have been scanned at 500 dpi and compressed using WSQ. There are a total of 4320 rolled fingers from 216 persons and all 10 fingers are scanned. Two impressions of each finger is available in addition to the plain finger images. All the images of a person are formatted as a NIST record.

- *NIST-30* (http://www.nist.gov/srd/nistsd30.htm): NIST-30 are dual resolution (500 dpi and 1000 dpi) fingerprint images to test fingerprint compression algorithms. The data consists of 36 ten-print paired cards with both the rolled and plain images scanned at dual resolution. The cards are segmented into the 10 rolled prints and the plain prints and stored in ANSI/NIST formatted files.

At the 2000 IEEE International Conference on Pattern Recognition, the first Fingerprint Verification Contest (FVC) was organized by the University of Bologna [100].

This competition and the followup one has resulted in the *FVC 2000 and FVC 2002 Fingerprint Databases*. As most of the NIST databases deal with rolled fingerprints involving scanned images, the FVC 2000 databases have been collected using live scan devices using three different scanners including low-resolution optical (256 dpi), CMOS (smaller scan area of $0.6" \times 0.8"$) and a full ($1" \times 1"$) gray scale optical FBI-certified scanner. In addition there is a synthetic database of fingerprints. Yet another difference of FVC 2000 database is that there are 8 samples per finger which can enable algorithms to understand the variations across matching fingers. However, the database consists of 100 fingers only and the same persons have not been imaged for all the scanners. Each database has a total of 800 images.

A newer version of the database is being planned in FVC 2002. More details about FVC 2000 and FVC 2002 are available at [12].

15.3.3 Speaker recognition databases

There are several public speech databases including the multi-modal database M2VTS described above that can be used for speaker recognition. There are several issues in designing a speaker recognition database including type of microphone used, the acoustic environment, type of speech (prompted, conversational or fixed), the duration of the acquired speech signal and the channel used. In addition, other parameters such as the number of speakers and the interval between sessions and number of sessions used is similar to other biometric databases. For an article on speaker databases, see [79].

Hence, there are simply many more parameters to voice databases than to finger or face databases.

- *NIST* (http://www.nist.gov/speech/tests/spk/index.htm): The original NIST speaker database is derived from Switch board I and Switch board II data sets. The data sets were collected by the Linguistic Data Consortium (LDC) and consists of 2,728 five-minute conversations involving 640 college student subjects as speakers.

  The subjects are allowed to make one call per day and every call by the same person involved use of a different hand set whereas the receiver always had the same hand set.

- *TIMIT and NTIMIT* (http://www.ldc.upenn.edu/Catalog/LDC93S2.html): The DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus (TIMIT) is one of the first databases in the speaker recognition area. The NYNEX NTIMIT Speech Corpus CD-ROMs (NTIMIT) database has been recreated by an artificial mouth talking into a hand set and recording after the signal was transmitted through long distance telephone lines. The database consists of 630 speakers where the subjects read sentences in a acoustic sound booth.

- *The YOHO Speaker Verification database* (http://www.ldc.upenn.edu/Catalog/LDC94S16.html): The YOHO data set is useful for text-dependent speaker recognition experiments. The data set consists of 128 subjects speaking the prompted digit phrases into high quality hand sets in an office environment. The data has been collected over 4 enrollment sessions and 10 verification sessions with at least several days interval between the sessions.

- *ELRA databases* (http://www.icp.grenet.fr/ELRA/). The European Language Resources Association (ELRA) has several speaker databases in non-English languages and non-native English speakers. The Speaker Identification and Verification Archives (SIVA) database is an Italian speaker database consisting of 671 subjects. The subjects speak the prompted words and digits and read text into a telephone handset in an office or home environment. The data has been collected over several session and the sessions separated by several days.

  The PolyVar is a speaker verification corpus comprised of native and non-native speakers of french, mainly from Switzerland. It consists of read and spontaneous speech in Swiss and French amounting to 160 hours of speech. Thirty-one speakers called from 2 to 10 times and 41 speakers made more than 10 calls.

  The POLYCOST database is collected under the European COST 250 project. Most of the speech is non-native English speakers with some speech in the speaker's native tongue covering 13 European countries. The speech were collected over ISDN lines. The impact of language in speaker recognition can be tested with this database.

- *OGI* (http://cslu.cse.ogi.edu/corpora/spkrec/): A dedicated speaker recognition database involving 100 subjects is available from the OGI School of Science & Engineering. The subjects call system 12 times over a 2 year period involving noisy and different telephone environments using different types of hand sets. The speakers provide different types of data to make the database vocabulary independent.

More details about these and other databases for speaker recognition can be found in [79].

15.3.4 Signature databases

- *The UNIPEN database* (http://hwr.nici.kun.nl/unipen/): On-line handwriting recognition addresses the problem of recognizing handwriting from data collected with a sensitive pad which provides discretized pen trajectory information. Contrarily to other pattern recognition fields, such as speech recognition and optical character recognition, no significant progresses have been made, in the past few years, in on-line handwriting recognition to make large corpora of training and test data publicly available, and no open competitions have been organized. There is more to know about the history of Unipen. As of 1999, the international Unipen Foundation was installed to safeguard the distribution of a large database of on-line handwritten samples, collected by a consortium of 40 companies and institutes.

- *CADIX data set* :

15.4 Certifications

Certified products and solutions in the biometrics area are not yet there. Once again fingerprint related solutions have been more advanced in this area because of the longer history. There are two FBI certification

programs in the fingerprint area. The first one deals with certification of optical fingerprint scanners and the second one deals with the certification of WSQ implementation for accuracy. The International Computer Security Association (ICSA, formerly NCSA) has a certification program in the biometrics area. Common criteria certification is becoming more popular in biometrics recently as vendors are busy developing the protection profiles.

- *FBI* certification of fingerprint scanner: (http://www.fbi.gov/hq/cjisd/iafis/efts70/appendixf.htm) For the FBI's IAFIS system, there are strict guidelines for fingerprint image quality adherence during capture, display and printing. The specifications are covered in the Appendix-F of Criminal Justice Information Services (CJIS) Electronic Fingerprint Transmission Specification (http://www.fbi.gov/hq/cjisd/iafis/efts70/cover.htm). Appendix G describes the requirements for accrediting current live- and card-scan equipment for fingerprint acquisition. A list of FBI certified scanners as having been tested and found to be in compliance with the FBI's Integrated Automated Fingerprint Identification System (IAFIS) Image Quality Specifications (IQS) is available at http://www.fbi.gov/hq/cjisd/iafis/cert.htm. Note that the list includes certified live scanners, card scanners and printers.

- *WSQ* certification The Wavelet Scalar Quantization (WSQ) fingerprint image Compression logarithm is the standard for the exchange of gray level fingerprint images. The WSQ Specification defines a class of encoders and a single decoder with sufficient generality to decode compressed image data produced by any compliant encoder. Part III of the WSQ specification contains the specific parameter values that must be implemented by encoders for certification. A WSQ decoder must implement the full range of functionality contained in the WSQ Specification including even and odd length filters. Compliance with the WSQ Specification is determined by comparing the output from the implementation under test with the output from a NIST double precision reference implementation. The comparison criteria and accuracy requirements are contained in the WSQ Specification. More guide lines for certification are available at http://www.itl.nist.gov/iad/894.03/fing/cert_gui.html.

- *ICSA* certification: The ICSA (International Computer Security Association) used to certify biometric authentication products however recently they have discontinued this effort.

- *Common Criterion*

  Common Criteria is a set of generic security requirements to aid in the specification of products, functional and assurances related to system security attributes involving security measures implemented in hardware, software and firmware. The scope of common criteria includes standards for acceptance of evaluations of security products performed by independent labs and address protection of information from unauthorized disclosure, modification, or loss of use. However, the common criterion does not address accreditation, cryptographic algorithm selection, physical security related to the hardware and legal implications of common criteria. The key concept of common criteria is "target of evaluation" also known as *TOE* identifies the part of the system which is the subject of evaluation. The three sets of parties involved are: consumers, developers and evaluators. Consumers and other parties can specify the security functionality of a product in terms of standard *protection profiles* and independently select the evaluation assurance level from a defined set of seven Evaluation Assurance Levels (EALs) *EAL1-EAL7*. The evaluators use the TOE security threats, objectives, requirements, security function specifications and the assurance measures collectively known as *Security Target*. Biometrics when used in security environments would gain from a common criterion framework by defining a

protection profile and assurance levels that a product might have received. A protection profile for medium robust environments involving biometrics for the Dept. of Defense is described in [90]. The Common Criteria has been adopted by the ISO as an international standard ISO 15408.

15.5 Inter-operability issues

Inter-operability of biometric systems will be a necessity in large scale deployments of biometric recognition systems. At the hardware level, as each sensor involves a different interface, the devices are incompatible within the same class of biometrics. For example, two USB ported fingerprint scanners are incompatible and the same driver cannot interface with them. However, when the devices involve video frame interface, standards such as VFW and RS-170-based frame grabbing are unaffected. Once the signal is acquired, the inter-operability at the template level is almost impossible for any technology as the individual representations vary significantly and quite often are proprietary and undisclosed. Hence the applications may have to re-enroll the users if there is a change in the vendor or hardware. Using known standard APIs make the change of the software easier, the industry is however very far removed from plug-and-play operation between sensors and vendors.

16 Discussion

This document discussed many aspects of biometric authentication. We have identified a number of issues in this document that we attempt to summarize as follows:

- *Biometric identification:* Though in this document we defined both verification and identification as the execution of an authentication protocol, the two authentication methods are very different. Identification involves a search against a database **N** of n identities. The database **N** holds n individuals, which are enrolled for some reason; these reasons can be civilian or criminal.

  Verification, on the other hand, is a match of one biometric with only one or maybe few biometric templates. In a positive identification problem with a given biometric, the false accept (FA) rate of the entire system is roughly n times the FA rate of the $1:1$ verification for that particular biometric. Hence for reliable biometric person identification, the biometric $1:1$ matcher has to run at extremely low false accept rates. Hence it is not recommended to incorporate biometric identification, if security is an issue.

  Biometric screening, or negative identification, is by definition also a $1:n$ search problem (where n is the number of entities registered in the database) if only a biometric is used as "credential." For these large (pure biometric) screening protocols it is desirable that the false positive (FP) rate is low, else too many subjects will be incorrectly identified as a member of the most-wanted database **N** for each database inquiry. Here the FP rate of the screening system (again, only using a biometric) is approximately linearly proportional to the FP rate of the $1:1$ biometric matcher, as for the positive identification problem.

- *Biometric system evaluation:* Procedures for evaluating biometric authentication systems are not very well defined. In general, technology evaluations are of two types: technology evaluations and scenario evaluations. A scenario evaluation is the testing of a biometric authentication where the complete end-to-end system is installed at some test facility. These test facilities are often at some national research lab commissioned to do the comparative testing.

  Technology evaluations are the type of evaluations that have been in use in the speech and document retrieval areas for quite a while now. Some organization, often a government agency, releases databases of test data at some point and test participants submit their algorithms within some period of time after the release of the test data.

  Both these evaluation methods rely on the selection of volunteers, which has unknown repercussions on the test results. Technology evaluations seem to be somewhat more robust to sample collection procedures and it seems that technology evaluations are a better way to compare the accuracy numbers of biometric matchers. Note that a technology evaluation in general measures more attributes of a system than just accuracy.

- *Realistic error estimates:* In marketing material of biometric systems, all too often error rates are not reported or poorly reported or, worse, the systems are claimed to be 100% accurate. In laboratory environments, one may achieve close to 100% accuracy when good quality biometric samples are used. For applications in the field, where the conditions are far from ideal, unavoidable noise and other ambient conditions will adversely affect the system performance.

  When reporting system accuracy, the accuracy should at least be reported as two numbers, the FA and FR rates; further an estimate of the Failure to Enroll (FTE) should be given. Along with these

numbers, some feel for the database quality should be given, e.g., at a minimum, it should be stated how the samples are obtained and some sample biometrics should be shown. Confidence intervals should be supplied to give an indication of the the significance of the estimates.

- *Enrollment:* This is perhaps the most ignored aspect of biometric authentication systems. Enrollment should be according to well specified policies (at least for enrollment in noncriminal databases). This is obviously true from the point of view of the public. It is even more true from a technical point of view; biometric sample collection during enrollment has to be conducted very carefully. The authentication system is only as accurate as the accuracy of the enrollment and this is important to keep in mind.

  Construction of screening databases **N** of course is largely done by federal agencies that maintain criminal databases.

- *User interface studies:* The second most ignored aspect of biometric authentication systems, and we probably have not stressed this enough in this document, is the user interface. Many existing biometric authentication systems are particularly awkward to interact with This is probably mostly the result of attempts to control the quality of the input biometric sample and ask the human to adapt to the input device. Problems with user friendliness may impact the performance of biometric authentication in terms of FA and FR, however there are no studies done by user interface researchers.

  The weakest link in any system design is where the user interacts with the system, biometric authentication systems are of course no exception. There is almost nothing known about the influences of the user interface on the system accuracy. Much work is done on this aspect at all the biometric authentication installations that are being put in place; unfortunately, this has not resulted in a comprehensive description of the state in the art in this area.

- *Security issues:* Security (safety) is the protection of a service or system against threats. There is the threat of a violent attack on the system. There is the threat of certain people impersonating other people (impersonate the users of the system or other individuals) to somehow violate or attack the application or the users of the application.

  The defense against such attacks could well become a convoluted process of security measures and countermeasures, as is the case today in message authentication in the area of computer security. One avenue is a protocol involving various authentication methods (combinations of possessions, knowledge or biometrics). Here biometric authentication systems can be inherently more secure than legacy authentication systems because there is, in theory, a "more secure" (non-repudiable) linking of individuals to the universally better accepted identity databases (passports, birth certificates).

  Many unanswered questions about *how* to make biometric authentication work without creating additional security loopholes still remain unanswered. Not much work in this area is being done. While there are the beginnings of Protection Profiles, in general there is not much mention of biometrics in the security literature.

- *Privacy issues:* The possibility of infringing on civil liberties in biometric authentication systems is very real. Integration of voluntary verification systems and involuntary screening systems is a concern beyond the traditional privacy concerns like having a biometric sample taken.

Biometrics is an area of many facets, many of which we have attempted to review, or at least to explain in this document. Our intent was to allow the question: "Which biometric is best?" to be answered in a

better-informed, detailed way using this document. While the answer may be only tentative, and surely not definitive, at least the question can be phrased a little better by asking: “Which biometric is best, given that the requirements for the application are specified as follows ...” Whether we have succeeded in this goal, we will of course not know without feedback from the readers. Therefore, the authors would like to invite you to contact them about any remaining unanswered concerns.

References

[1] AAMVA Standards Working Group. AAMVA Standard for the Driver License / Identification Card 2000. Technical Report AAMVA DL/ID-2000, The American Association of Motor Vehicles Administrators, June 2000.

[2] American National Standards Institute (ANSI). Biometric Information Management and Security. Technical Report X9.84-2001, http://www.x9.org/books.html, 2001.

[3] D. Ballard and C. Brown. *Computer Vision*. Prentice-Hall, Englewood Cliffs, NJ, 1982.

[4] H.S.M. Beigi, S.H. Maes, U.V. Chaudhari, and J.S. Sorensen. IBM model-based and frame-by-frame speaker recognition. In *Speaker Recognition and its Commercial and Forensic Appications*, Avignon, April 1998.

[5] P.H. Belhumeur, J.P. Hespanha, and D.J. Kriegman. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Transanctions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, July 1997.

[6] J. Bery. The history and development of fingerprinting. In H.C. Lee and R.E. Gaensslen, editors, *Advances in Fingerprint Technology*, pages 1–38. CRC Press, Boca Raton, FL, 1994.

[7] W. Bicz, Z. Gurnienny, and M. Pluta. Ultrsound sensor for fingerprints recognition. In *Proc. of SPIE, Vol. 2634, Optoelectronic and electronic sensors*, pages 104–111, June 1995.

[8] BioAPI Consortium. *BioAPI Specification Version 1.1*. The BioAPI Consortium, March 2001.

[9] Biometrics Consortium. http://www.biometrics.org.

[10] Biometrics Market Intelligence. Biometric API (BAPI), www.iosoftware.com/products/licensing/ bapi. 1(1):9, February 2002.

[11] Biometrics Working Group. Best practices in testing and reporting performance of biometric devices. http : //www.afb.org.uk/bwg/bestprac.html, 2000.

[12] Biometric Systems Lab, Pattern Recognition and Image Processing Laboratory, and U.S. National Biometric Test Center.

[13] Biometric Systems Lab. *HaSIS - A Hand Shape Identification System*.

[14] D.M. Blackburn, M. Bone, and P.J. Phillips. FRVT 2000: Facial recognition vendor test. Technical report, DoD Counterdrug Technology Development Office, Defence Advance Research Project Agency, National Institute of Justice, Dahlgren, VA; Crane, IN; Arlington, VA, December 2000.

[15] R.M. Bolle, J.H. Connell, N. Haas, R. Mohan, and G. Taubin. Veggievision: A produce recognition system. In *Proc. Third IEEE Workshop on Applications of Computer Vision*, pages 244–251, December 1996.

[16] R.M. Bolle, J.H. Connell, A. Hampapur, E. Karnin, R. Linsker, G.N. Ramaswamy, N.K. Ratha, A.W. Senior, J.L. Snowdon, and T.G. Zimmerman. Biometric technologies ... emerging into the mainstream. Technical Report RC22203 (W0110-041), IBM Research Division, Yorktown Heights, NY, October 2001.

[17] R.M. Bolle, J.H. Connell, S. Pankanti, and N. Ratha. On the security of biometrics authentication. IBM Technical Report, 2002.

[18] R.M. Bolle, J.H. Connell, and N.K. Ratha. Biometrics perils and patches. Patttern Recognition, *In print*, 2002.

[19] R. M. Bolle, N. K. Ratha, and S. Pankanti. Evaluating authentication systems using bootstrap confidence intervals. In *Proceedings of AutoID'99*, pages 9–13, Summit, NJ, USA, October 1999.

[20] R. M. Bolle, N. K. Ratha, and S. Pankanti. Evaluation techniques for biometrics-based authentication systems (FRR). In *Proc. 15th Int. Conf. on Pattern Recognition*, pages 835–841, September 2000.

[21] F.A. Bouchier, J.S. Ahrens, and G. Wells. Laboratory evaluation of the IriScan prototype biometric identifier. Technical Report SAND96-1033 RS-8232-2/960378, Sandia National Laboratories, Albuquerque, NM, April 1996.

[22] V. Bouletreau, N. Vincent, R. Sabourin, and H. Emptoz. Handwriting and signature: One or two personality identifiers? In *Proc. of the 14th International Conference on Pattern Recognition*, volume II, pages 1758–1760, Brisbane, Austria, August 1998.

[23] C.M. Brislawn, J.N. Bradley, R.J. Onyshczak, and T. Hopper. The FBI compression standard for digitized fingerprint images. In *Proc. of SPIE*, volume 2847, pages 344–355, August 1996.

[24] R.R. Brooks and S.S. Iyengar. *Multi-sensor Fusion: Fundamentals and Applications with Software*. Prentice-Hall, Upper Saddle River, New Jersey, 1997.

[25] R. Brunelli and T. Poggio. Face recognition: Features versus templates. *IEEE Transanctions on Pattern Analysis and Machine Intelligence*, 15(10):1042–1052, October 1993.

[26] J. Campbell. Speaker recognition. In A.K. Jain, R.M. Bolle, and S. Pankanti, editors, *Biometrics: Personal Identification in Networked Society*, pages 165–190. Kluwer Academic Press, Boston, 1999.

[27] R. Cappelli, A. Lumini, D. Maio, and D. Maltoni. Fingerprint classification by directional image partitioning. *IEEE Transanctions on Pattern Analysis and Machine Intelligence*, 21(5):402–421, May 1997.

[28] CBEFF Technical Development Team. Common Biometric Exchange File Format (CBEFF). Technical Report NISTIR 6529, The National Institute of Standards and Technology (NIST), January 2001.

[29] R. Chellappa, C.L. Wilson, and S. Sirohey. Human and machine recognition of faces: A survey. *Proceedings of the IEEE*, 83(5):705–740, May 1995.

[30] A. M. Choudhary and A. A. S. Awwal. Optical pattern recognition of fingerprints using distortion-invariant phase-only filter. In *Proc. of SPIE, Vol. 3805, Photonic devices and algorithms for computers*, pages 162–170, Oct. 99.

[31] R. Clarke. Human identification in information systems: Management challenges and public policy issues. *Information Technology & People*, 7(4):6–37, December 1994.

[32] J. Clark and A. Yuille. *Data Fusion for Sensory Information Processing Systems*. Kluwer Academic Publishers, Boston, MA, 1990.

[33] W.G. Cochran. *Sampling Techniques*. Wiley Series In Probability and Matchematical Statistics. John Wiley & Sons,, NewYork, 3 edition, 1977.

[34] Criminal Justice Information Services (CJIS). Electronic fingerrint transmission specification. Technical Report CJIS-RS-0010 (V7), Criminal Justice Information Services Division, Washington, D.C., January 1999.

[35] R. Curbelo. Noisy fingerprint identification by artificial neural networks. In *Proc. of SPIE, Vol. 3728, 9th Workshop on Virtual Intelligence/ Dynamic Neural Networks*, pages 432–449, March 99.

[36] Daubert Update. *Latent Print Examination, Fingerprints, Palmprints and Footprints*. http : //onin.com/fp/.

[37] J. G. Daugman and G. O. Williams. A proposed standard for biometric decidability. In *CardTechSecureTech*, pages 223–234, Atlanta, GA, 1996.

[38] J. G. Daugman. High confidence visual recognition of persons by a test of statistical independence. *IEEE Transanctions on Pattern Analysis and Machine Intelligence*, 15(11):1148–1161, Nov. 1993.

[39] J. Daugman. Recognizing persons by their itis patterm. In A.K. Jain, R.M. Bolle, and S. Pankanti, editors, *Biometrics: Personal Identification in Networked Society*, pages 103–122. Kluwer Academic Press, Boston, 1999.

[40] Digital Descriptor Systems, Inc. Non-contact fingerprint scanner. Technical report, http://www.ddsi-cpc.com/ productsmain.htm.

[41] G. Doddington, W. Liggett, A. Martin, M. Przybocki, and D. Reynolds. Sheep, goats, lambs and wolves: A statistical analysis of speaker performance. In *Proceedings of IC-SLD'98, NIST 1998 speaker recognition evaluation*, Sydney, Australia, November 1998.

[42] J.G.A. Dolfing. *Handwriting Recognition and Verification*. PhD thesis, University of Eindhoven, Eindhoven, the Netherlands, 1998.

[43] R. Donovan. *Trainable Speech Synthesis*. PhD thesis, Cambridge University, Engineering Department, Cambridge, U.K., 1996.

[44] C. Dorai, N. Ratha, and R.M. Bolle. Detecting dynamic behavior in compressed fingerprint videos: Distortion. In *Proc. IEEE Computer Vision and Pattern Recognition*, pages 320–326, June 2000.

[45] B. Duc, E.S. Bigün, J. Bigün, G. Maître, and S. Fischer. Fusion of audio and video information for multi modal person authentication. *Pattern Recognition Letters*, 18(9):835–843, 1997.

[46] A.L. Duwaer. *Data processing system with a touch screen and a digitizer tablet, both integrated in one input device*. US Patent No. 5231381, 1993.

[47] G.J. Edwards, C.J. Taylor, and T.F. Cootes. Interpreting faces using active appearance models. In *Third International Conference on Automatic Face and Gesture Recognition*, pages 300–305, Nara, Japan, April 1998.

[48] B. Efron. Bootstrap methods: Another look at the Jackknife. *Ann. Statistics*, 7:1–26, 1979.

[49] R.H. Ernst. *Hand ID system*. US Patent No. 3576537, 1971.

[50] B. Germain et al. Issues in large scale automatic biometric identification. In *IEEE Workshop on Automatic Identification Advanced Technologies*, pages 43–46, Stony Brook, NY, November 1996.

[51] FBI, U.S. Department of Justice, Washington, D.C. 20402. *The Science of Fingerprints, Classification and Uses*, 1984.

[52] Federal Bureau of Investigations. *WSQ gray-scale Fingerprint Image Compression Specification*, 1993.

[53] J. Ferryman, editor. *Performance Evaluation of Tracking and Surveillance*. 2002 IEEE Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii, December 2001.

[54] Fingerprint Data Interchange Workshop. Summary of the 1998 NIST Fingerprint Data Interchange Workshop, September 1998.

[55] D. T. Follette, E. B. Hultmark, and J. G. Jordan. *Direct optical input system for fingerprint verification*. IBM Technical Disclosure Bulletin: 04-74p3572, April 1974.

[56] S. Furui. Recent advances in speaker recognition. In Botgefors Bigun, Chollet, editor, *Audio- and Video-based Biometric Person Authentication*, volume 1206 of *Lecture Notes in Computer Science*, pages 237–252. Springer, 1997.

[57] R. Germain. Large scale systems. In A. Jain, R. Bolle, and S. Pankanti, editors, *Biometrics, Personal Identification in Networked Society*, pages 311–326. Kluwer Academic Publishers, Norwell, Mass., 1999.

[58] H.P. Graf. Sample-based synthesis of talking heads. In *Recognitio, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, pages 3–7, 2001.

[59] T. J. Grycewicz. Techniques to improve binary joint transform correlator performance for fingerprint recognition. *Optical Engineering*, 38(1):114–119, January 1999.

[60] Y. Hamamoto. A Gabor filter-based method for identification. In L. C. Jain, U. Halici, I. Hayishi, S. B. Lee, and S. Tsutsui, editors, *Intelligent biometric techniques in Fingerprint and face recognition*, pages 137–151. CRC Press, Boca Raton, 1999.

[61] L.P. Heck and M. Weintraub. Handset-dependent background models for robust text-independent speaker recognition. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, April 1997.

[62] R. Hill. Retina identification. In A.K. Jain, R.M. Bolle, and S. Pankanti, editors, *Biometrics: Personal Identification in Networked Society*, pages 123–142. Kluwer Academic Press, Boston, 1999.

[63] J. Holmes, L. Wright, and R. Maxwell. A performance evaluation of biometric identification devices. Technical Report SAND91-0278/UC-906, Sandia National Laboratories, Albuquerque, NM; Livermore, CA, June 1991.

[64] L. Hong and A. Jain. Multimodal biometrics. In A.K. Jain, R.M. Bolle, and S. Pankanti, editors, *Biometrics: Personal Identification in Networked Society*, pages 327–344. Kluwer Academic Press, Boston, 1999.

[65] R.A. Huber and A. Headrick. *Handwriting Identification: Facts and Fundamentals*. CRC Press LCC, Boca Raton, Florida, April 1999.

[66] Intel Corporation. *Intel^{RT} Common Data Security Architecture (CDSA)*, 2002.

[67] I/O Software Inc. *Biometric Application Programming Interface (BAPI)*. http://www.iosoftware.com/products/licensing/bapi/glossary.htm, 2002.

[68] D. K. Isenor and S. G. Zaky. Fingerprint identification using graph matching. *Pattern Recognition*, 19(2):113–122, 1986.

[69] I.H. Jacoby, A.J. Giordano, and W.H. Fioretti. *Personnel Identification Apparatus*. US Patent No. 3648240, 1972.

[70] A.K. Jain, R.M. Bolle, and S. Pankanti, editors. *Biometrics: Personal Identification in Networked Society*. Kluwer Academic Publishers, Boston, Mass., 1999.

[71] A.K. Jain, R.M. Bolle, and S. Pankanti. Introduction to biometrics (Chapter 1). In A.K. Jain, R.M. Bolle, and S. Pankanti, editors, *Biometrics: Personal identification in networked society*, pages 1–41. Kluwer Academic Publishers, Boston, Mass., 1999.

[72] A.K. Jain, L. Hong, and S. Pankanti. Biometrics identification. *Communications of the ACM*, 43(2):91–98, 2000.

[73] A.K. Jain, S. Prabhakar, and L. Hong. A multichannel approach to fingerprint classification. *IEEE Transanctions on Pattern Analysis and Machine Intelligence*, 21(4):348–359, April 1999.

[74] A.K. Jain, A. Ross, and S. Pankanti. A prototype hand geometry-based verification system. In *2nd IEEE International Conference on Audio- and Video-based Biometric Person Authentication*, pages 166–171, Washington D.C., March 1999.

[75] A. K. Jain, L.Hong, and R. M. Bolle. On-line fingerprint verification. *IEEE Transanctions on Pattern Analysis and Machine Intelligence*, 19(04):302–313, April 1997.

[76] A. K. Jain, S. Prabhakar, L. Hong, and S. Pankanti. FingerCode: a filterbank for fingerprint representation and matching. In *Proc. of the CVPR 99, vol. 2*, pages 187–193, 1999.

[77] A. Jain, L. Hong, S. Pankanti, and R. Bolle. On-line identity-authentication system using fingerprints. *Proceedings of the IEEE*, 85:1365–1388, September 1997.

[78] B. Javidi and J. Wang. Position-invariant two-dimensional image correlation using a one-dimensional space integrating optical processor: application to security verification. *Optical Engineering*, 35(9):2479–2486, Sept. 1996.

[79] J. P. Campbell Jr. and D. A. Reynolds. Corpora for the evaluation of speaker recognition systems. In *Proc. of IEEE ICASSP*, volume 2, pages 829–832, 1999.

[80] S. Jung, R. Thewes, T. Scheiter, K. F. Gooser, and W. Weber. A low-power and high-performance cmos fingerprint sensing and encoding architecture. *IEEE Journal of Solid-state Cicuits*, 34(7):978–984, July 1999.

[81] Justice Blackmun. *Daubert v. Merrell Dow Pharmaceuticals*. 113 S. Ct. 2786, 1993.

[82] T. Kanade. *Picture Processing System by Computer Complex and Recognition of Human Faces*. PhD thesis, Dept. of Information Science, Kyoto University, 1973.

[83] I. Kansala and P. Tikkanen. Security risk analysis of fingerprint based verification in pdas. In *Proc. IEEE AutoID 2002*, pages 76–82, Tarrytown, NY, USA, March 2002.

[84] K. Karhunen. Uber lineare Methoden in der Warscheinlichkeitsrechnung. In *Ann. Acad. Sci. Fennicae, ser A1, Math. Phys.*, volume 37, 1946.

[85] C. Kaufman, R. Perlman, and M. Spencer. *Network Security, Private Communication in a Public World*. Prentice Hall PTR, Upper Saddle River, NJ, 1995.

[86] S. King, H. Harrelson, and G. Tran. Testing iris and face recognition in a personnel identification application. In F.L. Podio and Dunn J.S, editors, *Proceedings of the Biometrics Consortium Conference*. NIST, US Department of Commerce, Cristal City, VA, February 2002.

[87] M. Kirby and L. Sirovich. Application of the Karhunen-Loeve procedure for the cgharacterization of human faces. *IEEE Trans. on Pattern Analyis and Machine Intelligence*, 12(1):103–108, January 1990.

[88] M. Kirby and L. Sirovich. Application of the Karhunen-Loève procedure for the characterization of human faces. *IEEE Transanctions on Pattern Analysis and Machine Intelligence*, 12(1):103–108, 1990.

[89] J. Kittler, Y.P. Li, J. Matas, and M.U. Ramos Sánchez. Lip-shape dependent face verification. In Josef Bigün, Gérard Chollet, and Gunilla Borgefors, editors, *Audio- and Video-based Biometric Person Authentication*, volume 1206 of *Lecture Notes in Computer Science*, pages 61–68. Springer, March 1997.

[90] A. Kong, A. Griffith, D. Rhude, G. Bacon, and S. Shahs. Department of Defence & Federal Biometric System Protection Profile for Medium Robustness Environments. Technical Report Draft Version 0.02, US Department of Defence, March 2002.

[91] J. Koolwaaij. *Automatic Speaker Verification in Telephony: A Probabilistic Approach*. PhD thesis, University of Nijmegen, Nijmegen, the Netherlands, September 2000.

[92] H.C. Lee and R.E. Gaensslen, editors. *Advances in Fingerprint Technology*. CRC Press, Boca Raton, FL, 1994.

[93] L.L. Lee, T. Berger, and E. Aviczer. Reliable on-line human signature verification systems. *IEEE Transanctions on Pattern Analysis and Machine Intelligence*, 18(6):643–647, June 1996.

[94] S-H. Lee, S-Y. Yi, and E-S. Kim. Fingerprint identification by use of volume holographic optical correlator. In *Proc. of SPIE, Vol. 3715, Optical Pattern Recognition*, pages 321–325, March 99.

[95] R.Y. Liu and K. Singh. Moving blocks Jackknife and Bootstrap capture weak dependence. In R. LePage and L. Billard, editors, *Exploring the Limits of the Bootstrap*, pages 225–248, New York, NY, 1992. John Wiley & Sons, Inc.

[96] M.M. Loève. *Probability Theory*. Van Nostrand, Princeton, NJ, 1955.

[97] J. Luettin, N.A. Thacker, and S.W. Beet. Speaker identification by lipreading. In *Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP'96)*, volume 1, pages 62–65, 1996.

[98] S.H. Maes, J. Navratil, and U.V. Chaudhari. Converstaional speech biometrics. In J. Liu and Y. Ye, editors, *E-Commerce agents. Marketplace Solutions, Security Issues, and Supply Demands*, pages 166–179. Springer-Verlag, Berlin Heidelberg, 2001.

[99] J-F Mainguet, M. Pegulu, and J. B. Harris. Fingerchip^{TM}: thermal imaging and finger sweeping in a silicon fingerprint sensor. In *Proc. of AutoID 99*, pages 91–94, October 99.

[100] D. Maio, D. Maltoni, R. Cappelli, J.L. Wayman, and A.K. Jain. Fvc2000: Fingerprint verification competition. *IEEE Transactions on Pattern Analyis and Machine Intelligence*, 24(3):402–412, 2002.

[101] D. Maio and D. Maltoni. Direct gray-scale minutiae detection in fingerprints. *IEEE Transanctions on Pattern Analysis and Machine Intelligence*, 19(1):27–40, January 1997.

[102] T. Mansfield, G. Kelly, D. Chandler, and J. Kane. Biometric product testing final report. Technical Report CESG Contract X92A/4009309, Centre for Mathematics and Scientific Computing, National Physics Laboratory, Middlesex, UK, March 2001.

[103] T. Mansfield and J. Wayman. Best practices in testing and reporting performance of biometric devices, For biometrics working group. Technical Report Issue 2 draft 9, Centre for Mathematics and Scientific Computing, National Physics Laboratory, Middlesex, UK, February 2002.

[104] S.J. McPhee, M.A. Papadakis, L.M. Tierney, and R. Gonzales, editors. *Current medical diagnosis and treatment*. Appleton and Lange, Stamford, CT, 1997.

[105] M. H. Metz, Z. A. Coleman, N. J. Phillips, and C. Flatow. Holographic optical element for compact fingerprint imaging system. In *Proc. of SPIE, Vol. 2659, Optical security and counterfeit deterrance techniques*, pages 141–151, 1996.

[106] B. Miller. Vital signs of identity. *IEEE Spectrum*, 31(2):22–30, 1994.

[107] B. Moayer and K.S. Fu. A syntactic approach to fingerprint pattern recognition. *Pattern Recognition*, 7:1–23, 1975.

[108] B. Moayer and K. S. Fu. A tree system approach for fingerprint pattern recognition. *IEEE Trans. on Computers*, C-25(3):262–274, 1976.

[109] R.T. Moore. Automatic fingerprint identification systems. In H.C. Lee and R.E. Gaensslen, editors, *Advances in fingerprint technology*, pages 163–191. CRC Press, Boca Raton, FL, 1994.

[110] M.E. Munich and P. Perona. Camera-based ID verification by signature tracking. In *Proceedings of the European Conference on Computer Vision*, pages 782–796, 1998.

[111] R.N. Nagel and A. Rosenfeld. Computer detection of freehand forgeries. *IEEE Transactions on Computers*, 26(9):895–905, September 1977.

[112] V.S. Nalwa. Automatic on-line signature verification. *Proceedings of the IEEE*, 85(2):215–239, February 1997.

[113] J. Navratil, U.V. Chaudhari, and G.N. Ramashamy. Speaker verification using target and background dependent linear transforms and muti-system fusion. In *Proc. EUROSPEECH 2001*, November 2001.

[114] L. O'Gorman. Seven issues with human authentication technologies. In *Proc. IEEE AutoID 2002*, pages 185–186, Tarrytown, NY, USA, March 2002.

[115] L. O'Gorman. Seven issues with human authentication technologies, presentation. In *Proc. IEEE AutoID 2002*, pages 185–186, Tarrytown, NY, USA, March 2002.

[116] R.D. Olsen. Identification of latent prints. In H.C. Lee and R.E. Gaensslen, editors, *Advances in fingerprint technology*, pages 163–191. CRC Press, Boca Raton, FL, 1994.

[117] S. Pankanti, S. Prabhakar, and A.K. Jain. On the individuality of fingerprints. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume I, pages 805–812, Kauai, Hawaii, December 2001.

[118] C.P. Pfleeger. *Security in Computing*. Prentice Hall PTR, Upper Saddle River, NJ, 1996.

[119] P. J. Phillips, H. Moon, P. J. Rauss, and S. A. Rizvi. The FERET September 1996 database and evaluation procedure. In J. Bigun, G. Chollet, and G. Botgefors, editors, *Proceedings of the First International Conference on Audio and Video-based Biometric Person Authentication*, volume Lecture Notes in Computer Science 1206. Springer, April 1997.

[120] R. Plamondon and G. Lorette. Automatic signature verification and writer identification – the state of the art. *Pattern Recognition*, 22(2):107–131, December 1988.

[121] R. Plamondon and G. Lorette. Automatic signature verification and writer identification — The state of the art. *Pattern Recognition*, 22(2):107–129, 1989.

[122] J.F. Porter. *On the 30 error criterion*. Unpublished but described in J.L.. Wayman, editor, *National Biometric Test Center Collected Works,* National Biometric Test Center, University of San Jose, CA, 1997-2000.

[123] F.J. Prokoski and R. Riedel. Infrared identification of faces and body parts. In A.K. Jain, R.M. Bolle, and S. Pankanti, editors, *Biometrics: Personal Identification in Networked Society*, pages 191–212. Kluwer Academic Press, Boston, 1999.

[124] M. Przybocki and A. Martin. *The 1999 NIST Speaker Recognition Evaluation Speaker Detection and Speaker Tracking*. EUROSPEECH 99 6th European Conference on Speech Communication and Technology, Budapest, Hungary, September 1999.

[125] G.N. Ramaswamy. Conversational biometrics: The future of personal identification. Technical Report White paper, IBM Research Division, Yorktown Heights, NY, September 2001.

[126] A. Ranalli. Fingerprint matching via spatial correlation with regional coherence. In *Proc. of SPIE, Vol. 2932, Human detection and positive identification: methods and technologies*, pages 161–167, Jan. 1997.

[127] C. V. K. Rao. *Pattern Recognition Techniques applied to fingerprints*. PhD thesis, Linkoping University, Sweden, 1977.

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

[129] N.K. Ratha, V.D. Pandit, R.M. Bolle, and V. Vaish. Robust fingerprint authentication using local structural similarity. In *Fifth IEEE Workshop on Applications of Computer Vision*, pages 29–34, December 2000.

[130] N. K. Ratha, S. Chen, and A. K. Jain. Adaptive flow orientation based texture extraction in finger print images. *Pattern Recognition*, 28(11):1657–1672, November 1995.

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

[132] N. K. Ratha, K. Karu, S. Chen, and A. K. Jain. A real-time matching system for large fingerprint database. *IEEE Transanctions on Pattern Analysis and Machine Intelligence*, 18(8):799–813, Aug. 1996.

[133] N. Ratha and R. Bolle. Smartcard based authentication. In A. Jain, R. Bolle, and S. Pankanti, editors, *Biometrics, Personal Identification in Networked Society*, pages 369–384. Kluwer Academic Publishers, Norwell, Mass., 1999.

[134] N. Ratha, J.H. Connell, and R.M. Bolle. An analysis of minutiae matching strength. In J. Bigun and F. Smeraldi, editors, *Proceedings 3rd IEEE International Conference on Audio- and Video-Based Biometric Person Authentication*, pages 223–228. Springer Verlag, Heidelberg Berlin, June 2001.

[135] D.A. Reynolds. Comparison of background normalization methods for text independent speaker verification. In *Proceedings of the European Conference on Speech Technology*, pages 963–966, Rhodes, 1995.

[136] D.A. Reynolds. The effects of handset variability on speaker recognition performance: Experiments on the Switchboard Corpus. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, May 1996.

[137] D. Roberge, C. Soutar, and B. V. K. Kumar. Optimal trade-off filter for the correlation of fingerprints. *Optical Engineering*, 38(1):108–113, January 1999.

[138] A.E. Rosenberg and S. Parthasarathy. Speaker background models for connected digit password speaker verification. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, pages 81–84, Atlanta, GA, 1996.

[139] E. Rosenberg, J. DeJong, C-H Lee, and B-H Juang anf F.K. Soong. The use of cohort normalized scores for speaker verification. In J. Ohala, editor, *Proceedings of the 1992 International Conference Spoken Language Processing*, volume 1, pages 599–602. University of Alberta, Alberta, CA, 1992.

[140] T. Rowley. Silicon fingerprint readers: A solid state approach to biometrics. In *Proc. of the CardTech/SecureTech, Orlando, Florida*, pages Vol. 1, 152–159, Washington D.C., May 1997.

[141] T. Ruggles. Comparison of biometric techniques. Technical report, The California State Legislature, http : //biometric-consulting.com/bio.htm, April 1996. Revised May 8, 2001.

[142] A. Samal and P.A. Iyengar. Automatic recognition and analysis of human faces and facial expressions: A survey. *Pattern Recognition*, 25(1):65–77, 1992.

[143] B. Schneier. The uses and abuses of biometrics. *Communications of the ACM*, 42(8):136, 1999.

[144] A. W. Senior. Recognizing faces in broadcast video. In *IEEE International Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, pages 105–110, September 1999.

[145] A. Senior. A combination fingerprint classifier. *IEEE Transactions on Pattern Analyis and Machine Intelligence*, 23(10):1165–1174, 2001.

[146] D. Setlak. Fingerprint sensor having spoof reduction features and related methods. US Patent Number: 5,953,441, September 1999.

[147] D.P. Sidlauskas. *3D hand profile identification apparatus*. US Patent No. 4736203, 1988.

[148] L. Sirovich and M. Kirby. Low-dimensional procedure for the characterization of human face. *J. Optical Society of America*, 4:519–524, 1987.

[149] M. K. Sparrow and J. Penelope. A topological approach to the matching of single fingerprints: development of algorithms for use on rolled impressions. Technical Report Special Publication 500-126, National Bureau of Standards, 1985.

[150] R.W. Sproat. *Multilingual Text-to-Speech Synthesis: The Bell Labs Approach,* Lucent Technologies Staff, Bell Laboratories, Lucent Technologies, Murray Hill, NJ, USA. Kluwer Academic Publishers, Boston, Mass., October 1997.

[151] S.N. Srihari, S-H Cha, H. Arora, and S. Lee. Individuality of handwriting: A validation study. In *International Conference on Document Analysis and Recognition*, pages 105–109, Seattle, WA, September 2001.

[152] W. Stallings. *Network and Internetwork Security*. Prentice Hall, Englewood Cliffs, 1995.

[153] A. Stoianov, C. Soutar, and A. Graham. High-speed fingerprint verification using an optical correlator. *Optical Engineering*, 38(1):99–107, January 1999.

[154] J. D. Stosz and L. A. Alyea. Automated system for fingerprint authentication using pores and ridge structures. In *Proc. of SPIE, Vol. 2277, Automatic system for the identification and inspection of humans*, pages 210–223, Oct. 94.

[155] D.L. Swets and J. Weng. Using discriminant eigenfeatures for image retrieval. *IEEE Transanctions on Pattern Analysis and Machine Intelligence*, 18(8):831–836, August 1996.

[156] C.J. Tilton. An emerging biometric standard. In S. Pankanti, R.M. Bolle, and Jain, editors, *IEEE Computer Magazine,* Special Issue on Biometrics, volume 1, pages 130–135. February 2001.

[157] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuro Science*, 3(1):71–86, 1991.

[158] UK Government Biometrics Working Group. Biometric Device Protection Profile (BDPP). Technical Report Draft Issue 0.28, September 2001.

[159] U.S. Department of Justice. *Solicitation: Forensic Friction Ridge (Fingerprint) Examination Validation Studies*. National Institute of Justice, Office of Science and Technology, Washington, D.C. 20531, March 2000.

[160] U.S. v. Byron Mitchell. *Criminal Action No. 96-407*. U.S. District Court for the Eastern District of Pennsylvania.

[161] Visionics Corporation, Inc. http://www.visionics.com/.

[162] R. Wang, T. J. Hua, J. Wang, and Y. J. Fan. Combining of Fourier transform and wavelet transform for fingerprint recognition. In *Proc. of SPIE, Vol. 2242, Wavelet Applications*, pages 260–270, March 94.

[163] C. I. Watson. *NIST special database 10: Supplemental Fingerprint Card Data for NIST Special Database 9*. Advanced Systems Division, Image Recognition Group , National Institute for Standards and Technology, February 1993.

[164] C. I. Watson. *NIST special database 14: Fingerprint Card Pairs 2*. Advanced Systems Division, Image Recognition Group , National Institute for Standards and Technology, February 1993.

[165] C. I. Watson. *NIST special database 4: 8-bit Gray scale Images of Fingerprint Image Groups*. Advanced Systems Division, Image Recognition Group , National Institute for Standards and Technology, February 1993.

[166] C. I. Watson. *NIST special database 9: Mated Fingerprint Card Pairs*. Advanced Systems Division, Image Recognition Group , National Institute for Standards and Technology, February 1993.

[167] C. I. Watson. *NIST special database 24: NIST Digital Video of Live-scan Fingerprint Database*. Advanced Systems Division, Image Recognition Group , National Institute for Standards and Technology, February 1998.

[168] J.L. Wayman. A scientific approach to evaluating biometric systems using mathematical methodology. In *Proceedings of CardTech/SecureTech.*, pages 477–492, Orlando, FL, May 1997.

[169] J.L. Wayman. *National Biometric Test Center Collected Works*. National Biometric Test Center, San Jose, CA, August 2000.

[170] J. H. Wegstein and J. F. Rafferty. Matching fingerprints by computer. Technical Report Technical note 466, National Bureau of Standards, 1969.

[171] D. Welsh and K. Sweitzer. *Presented at CardTech/SecureTech.* Orlando, FL, May 1997.

[172] R.P. Wildes, J.C. Asmuth, G.L. Green, S.C. Hsu, R.J. Kolczynski, J.R. Matey, and S.E. McBride. A machine-vision system for iris recognition. *Machine Vision and Applications*, 9:1–8, 1996.

[173] R.P. Wildes. Iris recognition: An emerging biometric technology. *Proceedings of the IEEE*, 85(9):1348–1363, September 1997.

[174] C. L. Wilson, C. I. Watson, and E. G. Paek. Combined optical and neural fingerprint matching. In *Proc. of SPIE, Vol. 3073, Optical Pattern Recognition*, pages 373–382, March 97.

[175] L. Wiskott and C. von der Malsburg. Recognizing faces by dynamic link matching. In *Proceedings of the International Conference on Artificial Neural Networks*, pages 347–352, 1995.

[176] T. Worthington, T. Chainer, J. Wilford, and S. Gunderson. IBM dynamic signature verification. *Computer Security*, pages 129–154, 1985.

[177] M. Y-S Yao, S. Pankanti, N. Haas, N. Ratha, and R.M. Bolle. Quantifying quality: A case study in fingerprints. In *Proc. IEEE AutoID 2002*, pages 126–131, Tarrytown, NY, USA, March 2002.

[178] J.R. Young and H.W. Hammon. Automatic palmprint verification study. Technical Report RADC-TR-81-161 Final Technical Report, Rome Air Development Center, June 1981.

[179] R. Zunkel. Hand geometry based authentication. In A.K. Jain, R.M. Bolle, and S. Pankanti, editors, *Biometrics: Personal Identification in Networked Society*, pages 87–102. Kluwer Academic Press, Boston, 1999.