Biometric Systems

Script from the Course

Christoph Busch

June 7, 2022

Contents

1 Introduction to Biometrics

Biometric applications have the primary purpose, to provide access control with a non-delegable authentication factor. These applications are more convenient for users of an IT system on the one hand and increase the security of access control on the other hand. This manuscript covers a some aspects discussed in the course *Biometric Systems* and will look at how these techniques work, and what their strengths is. We also take a look at the weaknesses to learn how to prevent bypassing a security system. The compliance of biometric systems with European data protection regulations is of particular importance, which is why we will work Privacy Enhancing Technologies (PET) for biometric methods.

1.1 Overview

The International Standardisation defines the term *biometrics* as follows: *automated recognition of individuals based on their behavioral and biological characteristics* [6][1]. Biometric methods thus analyze human behavior and/or their biological characteristics. The biological characteristics are categorized into anatomical characteristics, which are shaped by structures of the body, and physiological characteristics, which are characterized by body functions such as the voice. The process of biometric authentication provides a unique link between an individual (i.e. the natural person) and their identity, regardless of where that identity is stored.

For more than a hundred years, criminal investigators have been using fingerprints to catch suspects on the basis of evidence at the scene of the crime. Today, computers speed up identification both online and at the scene: Automated Fingerprint Identification Systems (AFIS) compare traces found at the scene of a crime with millions of stored fingerprint images in just a few seconds. Nowadays the police has already introduced in some cities a mobile AFIS, which enables investigators to initiate this comparison even over mobile communication networks such as UMTS or GSM. But in addition to fingerprints, facial and iris images or representation of the hand geometry can be used as means of identification in a biometric process. It is no longer just criminal investigation offices that apply these technologies many commercial access control systems are now using biometrics for identification purposes. Biometrics, which is understood as the automated recognition of individuals based on their behavioural and biological characteristics, on the one hand exploits the rich set of anatomical characteristics related to the structure of the body (finger pattern, iris pattern etc.). These characteristics can be measured more or less directly. On the other hand behavioural or physiological characteristics are related to the function of the body such as the written signature, the

[1] you can find the definition online under `http://www.christoph-busch.de/standards.html\#370103`

voice or the typing pattern on the keyboard. Those functions create a signal, where Biometrics observes the emitted body signal.

1.2 Biometric Recognition

A biometric recognition process requires that an individual (i.e. the natural person) is known by the system in advance (enrolment) to create the necessary reference data. This is done in the enrolment procedure. Biometric systems can either be designed as verification systems or as identification systems.

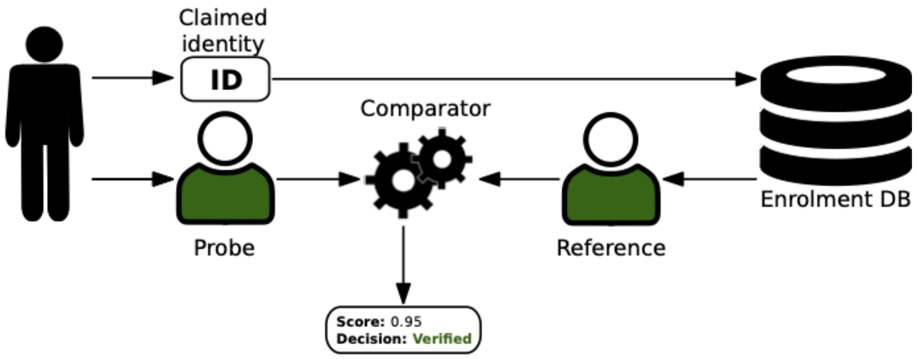


Figure 1.1: Biometric verification: confirming a biometric claim in a 1:1 comparison

In a verification system, the user specifies an identity to which - he claims - exists a reference in the system. If biometric systems are combined with an authentication document (e.g. a loyalty card), the biometric reference (e.g. a passport photo) may be stored on this document. At the time of verification, a comparison with exactly this one reference image is performed (1:1 comparison).

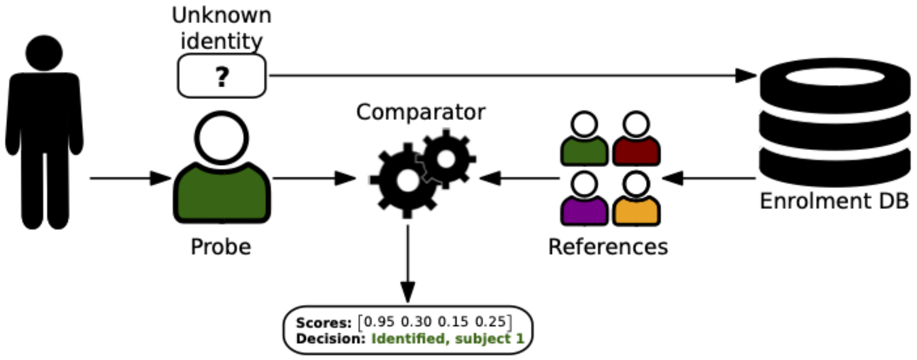

Unknown identity
?
Comparator
Enrolment DB
Probe
References
Scores: [0.95 0.30 0.15 0.25]
Decision: Identified, subject 1

Figure 1.2: Biometric identification: searching a reference in a 1:n search

In the case of an identification system, on the other hand, the captured image is compared with many images that have been enrolled, and the most similar reference record is determined from this set (1:n comparison). However, the similarity between two images must reach a pre-defined threshold, so that a reliable assignment of the identity associated with the most similar reference image can be done.

The biometric recognition process can be described with the following steps:

- **Acquisition:** Biometric characteristics are captured: a sensor, camera or other integrated capture device such as microphones observes the subject's characteristic and creates a digital representation, which is defined as the biometric sample (a scanned fingerprint, a digital portrait photo etc.)

- **Feature extraction:** This is a mathematical transformation applied to a biometric sample to derive distinguishing and repeatable numbers from the representation. These numbers are defined as biometric features, namely a concise representation of the original information. A biometric template is then understood to be a set of biometric features, which can be compared directly to biometric features from other presented biometric samples.

- **Enrolment:** In the enrolment process a biometric reference is created that is that one or more biometric samples or biometric templates are stored (in a database or in an ePass) and at the same time attributed to a subject. The reference can from then on be used for comparison.

- **Comparison:** This is a process in which probe sample stemming from the live biometric characteristic of one individual is compared against the biometric references of one or more individuals. The result of such comparison is a score that indicates the similarity (a value close but seldom identical to one) or dissimilarity (a value close to zero) of two samples.

Only after the comparison the recognition system is capable to decide on the comparison score, whether a presented sample matches or non-matches to a stored reference. The principle of biometric recognition is the same in all systems regardless of their particular technological design. In any case a biometric system must "learn" the biometric characteristic of the subject, before it can "recognize" an individual. Thus the information describing those biometric characteristics must be recorded and stored in data records. If a capture subject is asking for access authorization, the system compares current data with the data in the records. If they match to a specified level of certainty, the system recognizes the person and grants access.

The process steps are illustrated in figure 1.3, which provides a reference architecture for a biometric system [4].

1.2.1 Criteria for Biometric Characteristics

The procedure of enrolment, verification and identification are shown in Figure 1.3, from which also the essential components of a biometric system become apparent [4]. When constructing a biometric system, it is important to not only select appropriate technical components (i.e. capture device and embedded sensors, signal processing subsystem,

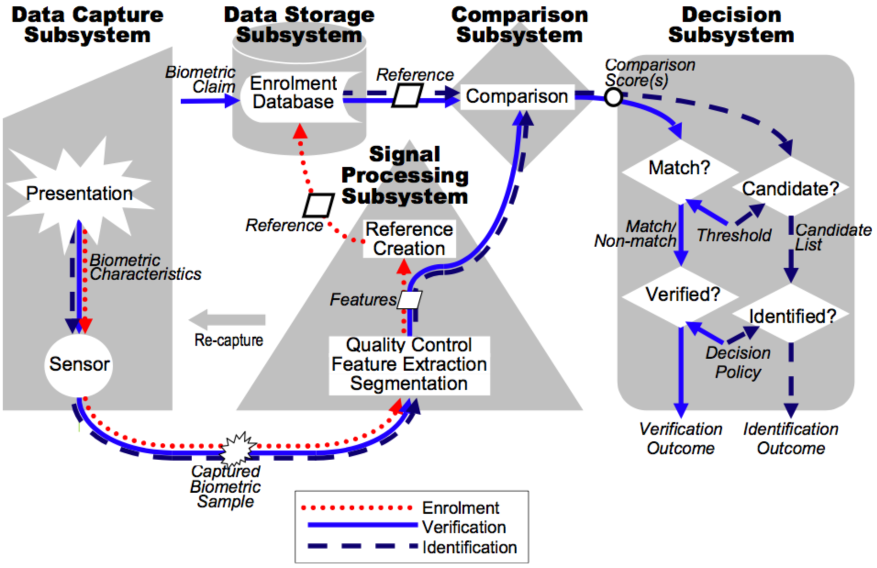
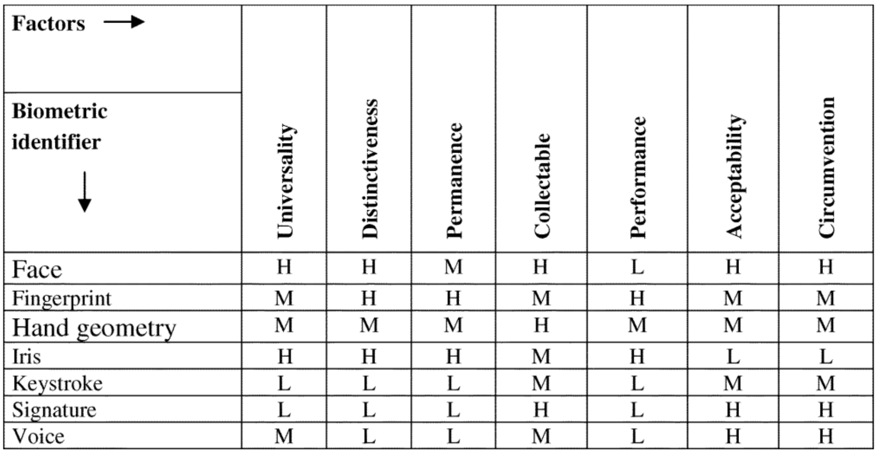


Figure 1.3: ISO/IEC reference architecture of a biometric system [4].

comparison subsystem, and decision subsystem) but also choose a suitable biometric characteristic for the target group that meets the following criteria:

- **Universality:** every individual should have it.
- **Uniqueness:** is the characteristic distinctive such that any two individuals are sufficiently different.
- **Performance:** does a recognition system based on this biometric characteristic provide a reasonable biometric performance (low errors). Furthermore this property is associated with the throughput time (how does it take to capture the biometric characteristic and to extract features from the captured sample.
- **Permanence:** the characteristic should be invariant over time and features extracted thereof should be persistent and not be mutable over time. The ageing of the individual should not affect the feature vector.
- **Collectability:** the characteristic is measurable and the quantitative result is reproducible.
- **Acceptability:** the capture process provides a convenient measurement at low cost and is considered unobtrusive for the data subjects.
- **Security:** intended impostor attacks are hard, as it is difficult to collect the biometric characteristic and replicate a fake biometric characteristic, which would be capable to fool a sensor. Thus the security of a captures device measuring this biometric characteristic can be considered high.



Factors → Biometric identifier ↓	Universality	Distinctiveness	Permanence	Collectable	Performance	Acceptability	Circumvention
Face	H	H	M	H	L	H	H
Fingerprint	M	H	H	M	H	M	M
Hand geometry	M	M	M	H	M	M	M
Iris	H	H	H	M	H	L	L
Keystroke	L	L	L	M	L	M	M
Signature	L	L	L	H	L	H	H
Voice	M	L	L	M	L	H	H

Figure 1.4: Comparison of biometric modalities [9].

If individual criteria are not met in a mono-modal system, multi-modal systems can be a solution. For example, a face recognition is combined with an iris recognition in order to achieve a sufficient recognition performance of the biometric system.

Biometric characteristics always arise under the influence of genes received from the parents. The face often resembles the face of the parents. Also, a certain behavior (e.g. the gait, the way of speaking) is often taken over from the parents or has been adapted from role models in the course of life. To a certain extent, characteristics are therefore genetically determined or characterized by behavior and the environment. However, a good and suitable biometric characteristic is primarily characterized by random factors. This property applies, for example, to the formation of a fingerprint or the pattern of the iris, which occur randomly during pregnancy. These patterns remain consistent and do not change when developing from child to adult.

1.2.2 Categories of Biometric Methods

In addition to the criteria of the biometric characteristics, we can also look at the categories for biometrics methods and then sort or benchmark biometric technology. The consideration of said categories makes it possible to evaluate certain products in a technology selection procedure. These are some examples of categories:

- **static / dynamic** - is a characteristic captured as a snapshot (e.g. photo of the face), or is a continuous measurement of a signal (e.g. recording of the voice or keyboard interaction) necessary.

- **cooperative / non-cooperative** - this category, which is for simplicity often referred to as *active / passive*, distinguishes between those methods of measurement in which the individual (i.e. *biometric capture subject*[2]) is aware that he/she

[2] biometric capture subject is defined `www.christoph-busch.de/standards.html\#370703`

interacts with a capture device (for instance the fingerprint sensor) versus such applications in which the affected individual is not aware of the data acquisition (e.g. video surveillance)

- **contact-free / contact-based** - for some modalities (such as fingerprints) the recording can be both contactless, i.e. using a digital photo, or contact-based, i.e. via a dedicated fingerprint sensor. The choice of the capture device has influence on any potential deformation of the finger (e.g. by the pressure applied during placement) and also on the acceptance of the procedure with regards to the subjects's concern about the transmission of diseases.

- **open / closed** - large biometric systems, such as the forensic applications of law enforcement agencies, are usually open so that data can be exchanged between different departments in a uniform standardized format. Even today's border control applications such as the one *EasyPASS*[3] at German Airports are open systems because identity documents which may have been produced outside of Germany must be able to be read when entering (see more in the chapter on **??**). For this, storing biometric data in the passport in a standard format [1, 2] is a mandatory prerequisite. On the contrary an operator that wants to ensure access control through biometrics can use a closed and *proprietary* storage format, but with a high risk: if the system vendor leaves the market, the reference data must be recoded or, if necessary, re-coded or recorded again.

- **supervised / unsupervised** - some biometric systems such as border control are deliberately operated under supervision only. For some biometric capture devices good robustness against attacks can be assumed, i.e. a sensor is not deceived by artefacts (i.e. counterfeit or plagiarized characteristics) [5]. Such capture devices are suitable for unattended systems in the physical access control to buildings, which can mean significant potential for savings in terms of staff. An online banking system with biometrics is another example of an unsupported application with high relevance.

- **positive identification / negative identification** - in the case of positive identification, the biometric statement consists of the statement *"I have a reference record in the company's database because I am an employee"*. In the case of negative identification, the biometric statement consists of the statement *"I have no reference record in the criminal records database because I am not a criminal"*.

- **environment sensitive** - performance of biometric techniques may be severely impacted by environmental factors. A facial recognition system is difficult to operate in direct sunlight whereas a speaker recognition system faces difficulties in the vicinity of a busy street.

[3] The EasyPASS application is a Biometric Border control: `https://www.easypass.de/EasyPass/DE/Was_ist_EasyPass/home_node.html`

In addition to the criteria discussed above, there are other obvious factors such as cost of procurement and cost and effort of operating the biometric system. Very relevant is also the question of familiarization of the affected person to the interaction with the capture device. A facial recognition photo can be taken without much instruction. If necessary, the supervisor must be trained, so that attention is paid to good lighting conditions. Training of the capture subject themself, however, is necessary for example for a signature recognition system. The “blind” writing on a tablet PC takes getting used to, and only over time a largely error-free interaction will be possible. From a technical point of view, it is particularly important for us to observe errors and to measure the recognition performance, which we will explore in chapter 2.

Strengths of Biometric Authentication

What are the strengths of biometric authentication? The classical authentication mechanisms, such as the knowledge authentication (password), authentication via tokens (keys) or the like are provided with distinct disadvantages. You can usually pass on your password and token in violation of a security guideline; you can forget about it or lose it. To prevent loss in the increasing number of logical and physical access controls, inappropriate storage locations or identical passwords are regularly used. In contrast, we cannot forget biometric characteristics and we cannot delegate them. Biometric applications allow identification of an individuals’s identity in logical and physical access control, and biometrics can solve problems of other authentication methods. Furthermore, in biometric authentication equality of security across different users prevails, as opposed to, for example, knowledge-based authentication in which “strong” or “weak” passwords can be chosen.

Weaknesses of Biometric Authentication

Biometric methods are usually used to either improve the usability of a technical system (e.g. fingerprint authentication on the iPhone or Samsung smartphone) or to improve the security of a technical system. However, it must be taken into account that the introduction of a biometric user recognition potentially introduces new security risks. Many attacks can be intercepted by cryptographic protocols and secure transmission of biometric data (see chapter **??**).

1.3 Vocabulary

Literature and science specifically in a multi-disciplinary community as in biometrics tends to struggle with a clear and non-contradictonary use and understanding of its terms. Thus ISO/IEC has undertaken significant efforts to develop a Harmonized Biometric Vocabulary (HBV) [6] that contains terms and definitions useful also in the context of discussions about presentation attacks. Without going into detail of the terminology definition process it is important to note that biometric *concepts* are always discussed in context (e.g. of one or multiple biometric subsystems) before a *term* and its *definition* for said concept can be developed. Thus terms are defined in groups and overlap of groups (”concept clusters”) and the interdependencies of its group members necessarily lead to revision of previously found definitions. The result of this work is published as ISO/IEC 2382-37:2017 [6]

The following list contains definitions of interest:

- **biometric characteristic** : biological and behavioural characteristic of an individual from which distinguishing, repeatable biometric features can be extracted for the purpose of biometric recognition (37.01.02)
- **biometric feature** : numbers or labels extracted from biometric samples and used for comparison (37.03.11)
- **biometric capture subject** : individual who is the subject of a biometric capture process (37.07.03)
- **biometric capture process** : collecting or attempting to collect a signal(s) from a biometric characteristic, or a representation(s) of a biometric characteristic(s,) and converting the signal(s) to a captured biometric sample set (37.05.02)
- **impostor** : subversive biometric capture subject who attempts to being matched to someone else’s biometric reference (37.07.13)
- **identity concealer** : subversive biometric capture subject who attempts to avoid being matched to their own biometric reference (37.07.12)

The use of biometric terms in a term paper **must** be compliant with the ISO/IEC SC37 Biometric Harmonized Vocabulary (ISO/IEC 2382-37:2012). In consequence replace for instance any occurrence of the term *matching* with *comparison* and use the term *template* only in a context, where you actually refer to a set of extracted biometric features.

1.4 Reading

Complementary reading for an introduction on biometrics is the overview provided with [8]. A tutorial on biometrics has been published as technical report of ISO/IEC JTC1 SC37 [**?**]. A general biometric system architecture is described in the ISO/IEC SC37 Standing Document 11[4]. The ISO/IEC SC37 Standardised Vocabulary [6] provides a Harmonized Biometric Vocabulary, which is available online version at:
`http://www.christoph-busch.de/standards.html`

2 Biometric Performance

2.1 Overview

- Metrics (FTC,FTX,FTE,FMR, FNMR, FAR, FRR, ROC, DET)
- confidence of measured error rates

2.2 Basic Metrics

Biometric performance is based on measurements conducted on a set of samples, which are available in a corpus or gallery. This corpus may contain a set of image samples, voice samples or other representation of a biometric characteristic. The fundamental metrics in a biometric system is the comparison score $c(Q, R)$ that is generated, when the algorithm under test compares a biometric probe or query Q with a biometric reference R. According to the ISO/IEC Harmonized Biometric Vocabulary [6] a comparison score is defined as:

Definition 1 (comparison score) *numerical value (or set of values) resulting from a comparison [6].*

Note that the term *matching score* is deprecated by ISO/IEC 2382-37 [6]. The comparison score can be expressed by a similarity score $s(Q, R)$ or distance score $d(Q, R)$

Definition 2 (similarity score) *comparison score that increases with similarity [6].*

Definition 3 (distance score / similarity score) *comparison score that decreases with similarity [6].*

For the distance score d, we expect the following properties to hold.

$$d(Q, R) \geq 0 \tag{2.1}$$

and

$$d(R, R) = 0 \tag{2.2}$$

The later can be considered as a self-comparison, e.g. two templates potentially derived with two different algorithms from one and the same sample should have a distance of zero or at least close to zero. The distance score d can be converted to a similarity score s using a monotonically decreasing function f. Examples for common conversions are the following

$$s = -d \tag{2.3}$$

$$s = -log(d) \tag{2.4}$$

$$s = \frac{1}{d} \tag{2.5}$$

For the case that the feature vectors are represented in binary form then the dissimilarity between the two vectors is expressed by *Hamming Distance* that essentially counts the number of different bits.

2.2.1 Distance Score

For a n-dimensional metric space an example for a distance metric is the *P-norm*, which is also known as *Minkowski metric.* In Figure 2.1 a two dimensional feature space with a probe feature vector Q and a reference feature vector R are illustrated. For the given example the two vectors are represented by

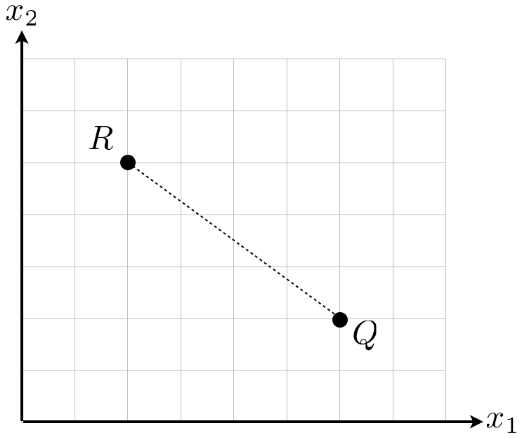


Figure 2.1: 2D feature space with probe Q and reference R

$$Q = \begin{pmatrix} q_1 \\ q_2 \end{pmatrix} = \begin{pmatrix} 6 \\ 2 \end{pmatrix}, \quad \text{and } R = \begin{pmatrix} r_1 \\ r_2 \end{pmatrix} = \begin{pmatrix} 2 \\ 5 \end{pmatrix}$$

As distance measure between those two points we express the length of the difference vector X as P-norm

$$||X||_p := \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}} \tag{2.6}$$

where n is the dimensionality of the feature space and p real number and $p \geq 1$.

City Block Norm For $p = 1$ the P-Norm simplifies to the the city block norm that is also called taxicab norm.

$$||X||_1 := \sum_{i=1}^{n} |x_i| \tag{2.7}$$

For the example from Figure 2.1 we derive

$$\begin{aligned}\sum_{i=1}^{n}|x_i| = \sum_{i=1}^{2}|q_i - r_i| &= |q_1 - r_1| + |q_2 - r_2| \\ &= |6-2| + |5-2| \\ &= 4+3 \\ &= 7\end{aligned}$$

Euclidian Nom For $p = 2$ the P-Norm simplifies to the Euclidian norm.

$$||X||_2 := \sqrt{\sum_{i=1}^{n}|x_i|^2} \tag{2.8}$$

For the example from Figure 2.1 we derive

$$\begin{aligned}\sqrt{\sum_{i=1}^{n}|x_i|^2} = \sqrt{\sum_{i=1}^{2}|q_i - r_i|} &= \sqrt{|q_1 - r_1|^2 + |q_2 - r_2|^2} \\ &= \sqrt{|6-2|^2 + |5-2|^2} \\ &= \sqrt{4^2 + 3^2} \\ &= \sqrt{25} \\ &= 5\end{aligned}$$

Maximum Norm For $p = \infty$ we get the infinity norm or maximum norm.

$$||X||_\infty := \sqrt{\sum_{i=1}^{n}|x_i|^\infty} = max\,(|x_1|, |x_2|, \cdots, |x_n|) \tag{2.9}$$

For the example from Figure 2.1 we derive

$$\begin{aligned}max\,(|x_1|, |x_2|, \cdots, |x_n|) = max\,(|x_1|, |x_2|) &= max\,(|q_1 - r_1|, |q_2 - r_2|) \\ &= max\,(|6-2|, |5-2|) \\ &= max\,(4, 3) \\ &= 4\end{aligned}$$

2.3 Biometric Failures

There are multiple failure associated with a acquisition of a biometric sample or with its processing. In Sections 2.3.1 to 2.3.3 we will discuss the failures that are associated with the deficiency of a biometric system to create a biometric reference for a data subject and subsequently in Sections 2.6 to 2.7 will consider errors that are attributed to biometric verification and identification systems.

2.3.1 Failure-to-Capture

A Failure-to-Capture Rate (FTC) is constituted, when the capture process could not generate a biometric sample of sufficient quality. This can be caused due to one of the following reasons:

1. The sample is not generated, as the characteristic is not placed properly on the capture device (e.g finger not covering the sensor area)
2. The captured signal is rejected by the automatic sample quality control algorithm.
3. The captured signal is stored as file, but rejected by the operator (staff expert) subsequent to visual inspection as it is not of sufficient quality

The ISO-definition [6] for the Failure-to-Capture-Rate (FTCR) is given by:

Failure-to-Capture Rate: *proportion of failures of the biometric capture process to produce a captured biometric sample of the biometric characteristic of interest.*

To estimate the FTCR we use the following formula:

$$FTCR = \frac{N_{tca} + N_{nsq}}{N_{tot}} \tag{2.10}$$

where N_{tca} is the number of terminated capture attempts, N_{nsq} is the number of images created with insufficient sample quality and N_{tot} is the total number of capture attempts. In consequence of a Failure-to-Capture are new capture attempt is initiated. This is illustrated in figure 2.2 .

2.3.2 Failure-to-eXtract

A Failure-to-eXtract is constituted, when the feature extraction process was not able to generate a biometric template. This can be caused due to one of the following reasons:

1. The algorithm itself declares that it cannot create a template from the input sample. This could be caused by a insufficient number of features that were identified e.g. only five minutia could be extracted from a fingerprint image.
2. Processing time of feature extraction algorithm exceeds the specified limit and thus the feature extraction is terminated
3. The feature extraction algorithm might suddenly crash during processing. In this case, some actions will be undertaken (e.g. start over application, repeat process, etc.) but if the crash happens all the time with the same sample then for this image a failure to extract feature will be constituted. There is currently no ISO-definition for the Failure-to-eXtract Rate.

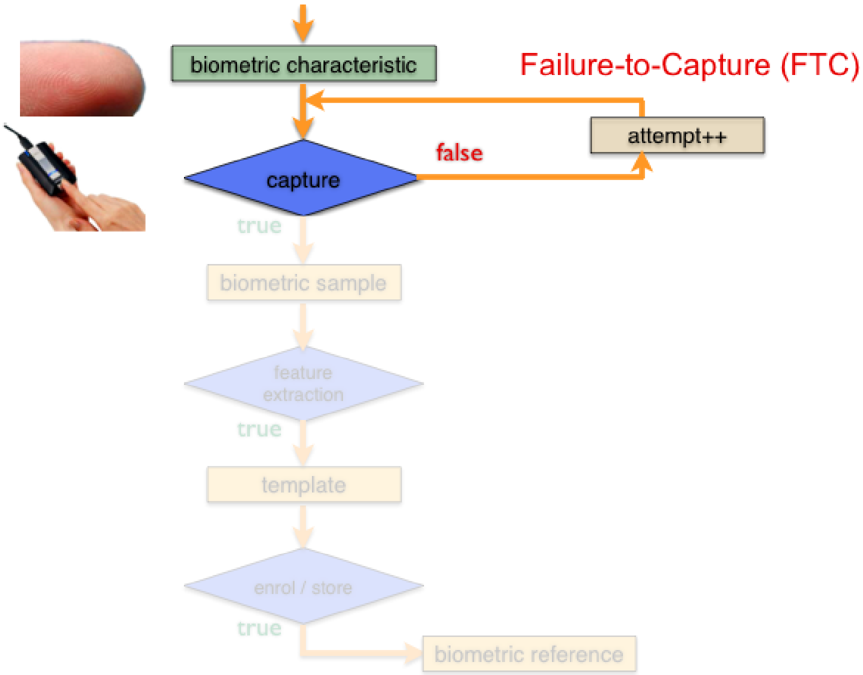


Figure 2.2: Failure-to-Capture (FTC)

To estimate the Failure-to-eXtract Rate (FTXR) we use the following formula:

$$FTXR = \frac{N_{ngt}}{N_{sub}} \tag{2.11}$$

where N_{ngt} is the number of cases, where no template was generated and and N_{sub} is the total number of biometric samples being submitted to the feature extraction component (i.e. the template generator). In an operational scenario the consequence of a Failure-to-eXtract is a new attempt including a new biometric sample creation and it subsequent processing. This is illustrated in figure 2.3 .

2.3.3 Failure-to-Enrol

A Failure-to-Enrol is constituted, when the biometric system is not capable to create for data subject a biometric reference. Thus the Failure-to-Enrol Rate (FTER) expresses the proportion of the population, for which the system fails to complete the enrolment process. This can be caused due to one of the following reasons:

1. The biometric characteristic of the subject (e.g. its fingerprint images) can not be captured at all.

2. For each evaluation setting, and if required instances of the same characteristic (e.g. left index finger instead right index finger) it is not possible to create for this

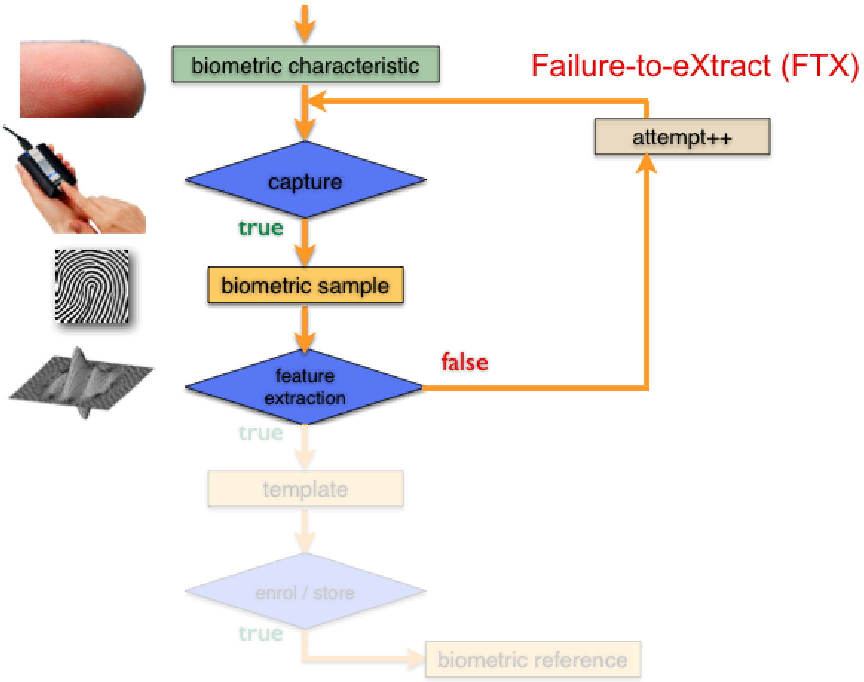


Figure 2.3: Failure-to-eXtract (FTX)

subject a template of sufficient quality (e.g. a feature set with minimum number of minutia)

The ISO-definition [6] for the Failure-to-Enrol-Rate (FTER) is given by:

Failure-to-Enrol Rate (ISO/IEC 2382-37): *proportion of a specified set of biometric enrolment transactions that resulted in a failureto create and store a biometric enrolment data record.*

To estimate the FTER we use the following formula:

$$FTER = \frac{N_{nec}}{N} \tag{2.12}$$

where N_{nec} is the number of cases, where we meet one of the two Failure-to-Enrol criteria and N is the total number of subjects, intended to be enroled in the biometric application. The consequence of a Failure-to-Enrol In an operational scenario is that for the capture subject a fallback procedure must be activated that should treat the individual in a non-discriminatory manner. The Failure-to-Enrol is illustrated in figure 2.4 .

2.3.4 Failure-to-Acquire

The Failure-to-Acquire Rate (FTAR) is essential for the verification process and estimates the likelihood that biometric comparison can not be completed due to potential

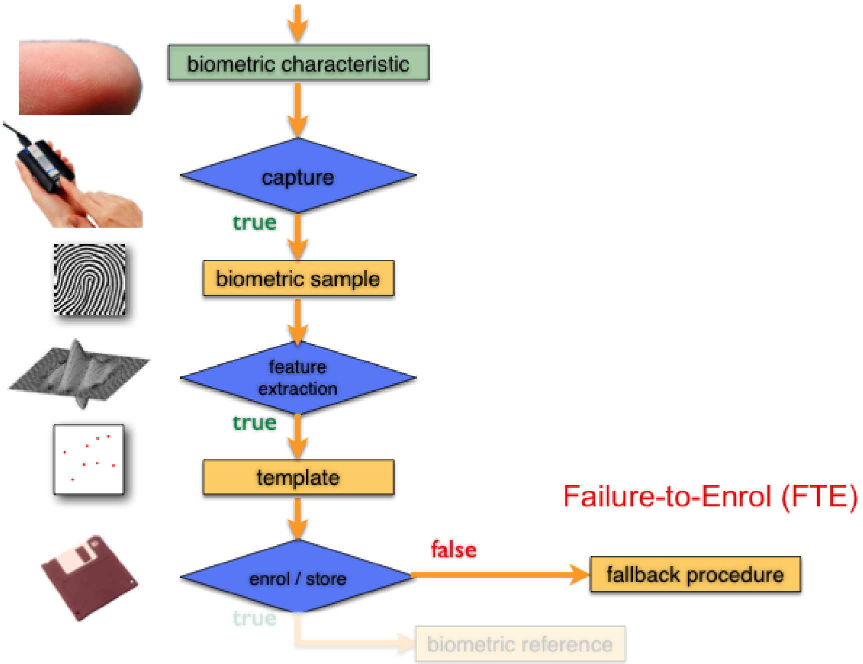


Figure 2.4: Failure-to-Enrol (FTE)

deficiencies in the live sample that is submitted as a probe. If there is no feature vector that can be compared to a biometric reference this can be caused due to one of the following reasons:

1. The is no biometric sample generated, which is expressed by the FTCR.
2. The feature extraction component failed to extract features as the number and/or quality of extracted features is not sufficient. This is expressed by the FTXR.

The ISO-definition [6] for the Failure-to-Acquire Rate (FTAR) is given by:

Failure-to-Acquire Rate (ISO/IEC 2382-37): *proportion of a specified set of biometric acquisition processes that were failure to accept for subsequent comparison the output of a data capture process.*

Note that in ISO/IEC 2382-37 a *probe* is defined as *biometric data input to an algorithm for comparison to a biometric reference(s).*To estimate the Failure-to-Acquire Rate we use the following formula:

$$FTAR = FTCR + FTXR * (1 - FTCR) \tag{2.13}$$

2.4 Performance measures

Given a corpus of biometric samples the following measures provide insight in the recognition accuracy of a feature extractor and biometric comparison subsystem. When computing the measures it should be considered to achieve independent trials. This is relevant, when for example multiple instances of a biometric characteristic are captured (e.g. 10 fingerprints or 2 eyes per data subject). The evaluator should respect that within-individual comparisons are not equivalent to between-individual comparisons, and shall thus not be included in the set of non-mated comparison trials.

2.4.1 False-Match

For non-mated (i.e. impostor) comparisons a False-Match constitutes the undesired case that a probe is matching a non-mated biometric reference, which has not been created for himself. The ISO-definition for the corresponding False-Match-Rate (FMR) is [7]:

False-Match-Rate (ISO/IEC 19795-1): *proportion of the completed biometric non-mated comparison trials that result in a false match.*

$$FMR(t) = \int_t^1 \phi_{nm}(s)ds \tag{2.14}$$

Together with the False-Non-Match-Rate (FNMR) the FMR is the key metric to be used in biometric technology testing and is understood to characterize a security property of a biometric system. Note that some literature is using the term False-Accept-Rate in the meaning of FMR.

2.4.2 False-Non-Match

For mated (i.e. genuine) comparisons a False-Non-Match constitutes the undesired case that a probe is not matching to a mated biometric reference, which has been created for the same subject from the same source (e.g. same index finger). The ISO-definition for the corresponding for False-Non-Match-Rate (FNMR) is [7]:

False-Non-Match-Rate (ISO/IEC 19795-1): *proportion of the completed biometric mated comparison trials that result in a false non-match.*

$$FNMR(t) = \int_0^t \phi_m(s)ds \tag{2.15}$$

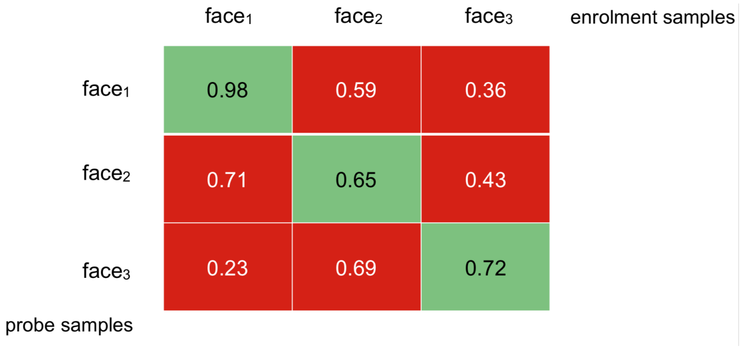
Note that in computing the FNMR we will count the Genuine-Match-Rate (GMR) first. The relationship between GMR and FNMR is as follows:

$$FNMR(t) = 1 - GMR(t) \tag{2.16}$$

Together with the False-Match-Rate (FMR) the FNMR is the key metric to be used in biometric technology testing and is understood to characterize a convenience property of a biometric system. Note that some literature is using the term False-Reject-Rate in the meaning of FNMR.

2.4.3 Similarity Matrix

In order to compute the above performance measures we need first to analyze the similarity scores achieved by our comparison subsystem. Figure 2.5 indicates the similarity scores for a face recognition subsystem for 3 subjects and 3 instances. The gallery was capture in 2 session, such that we have one enrolment sample and one probe sample per subject.



probe samples	$face_1$	$face_2$	$face_3$	enrolment samples
$face_1$	0.98	0.59	0.36	
$face_2$	0.71	0.65	0.43	
$face_3$	0.23	0.69	0.72	

Figure 2.5: Similarity matrix for 3 subjects and 3 instances

Similarity scores in green represent genuine scores (i.e. stemming from mated comparison trials) and similarity scores in red represent impostor scores (i.e. stemming from non-mated comparison trials).

If, while considering the recommendations regarding independent trials, multiple instances per subject are observed and evaluated, then the similarity matrix becomes more complex. In Figure 2.6 we can observe the similarity matrix for an evaluation, where we have multiple instances per subject and in addition multiple samples per instance captured. The number of similarity scores and impostor scores is significantly increasing.

In this case we have N instances (e.g. 10 fingers) and U samples captured per instance, which should stem from session, which are separated by at least one week. Figure 2.6 indicates that in total we have N*U=M samples, which can be used as either reference or probe sample. The size of the similarity matrix is M*M. For such a similarity matrix it becomes obvious that the number of impostor scores is significantly larger than the number of genuine scores. Counting the number of genuine scores, we have U samples (per instance) but from the possible U^2 we have to subtract U self-comparisons. Counting the number of impostor scores, we have $(M-1)*M$ comparison scores to consider.

A concrete example for the size of the similarity matrix: For a fingerprint system evaluation we invite 150 data subjects and record all 10 instances in 12 sessions. Our similarity matrix will then be of size

- Number of data subjects S: 150
- Number of biometric instances per subject: 10

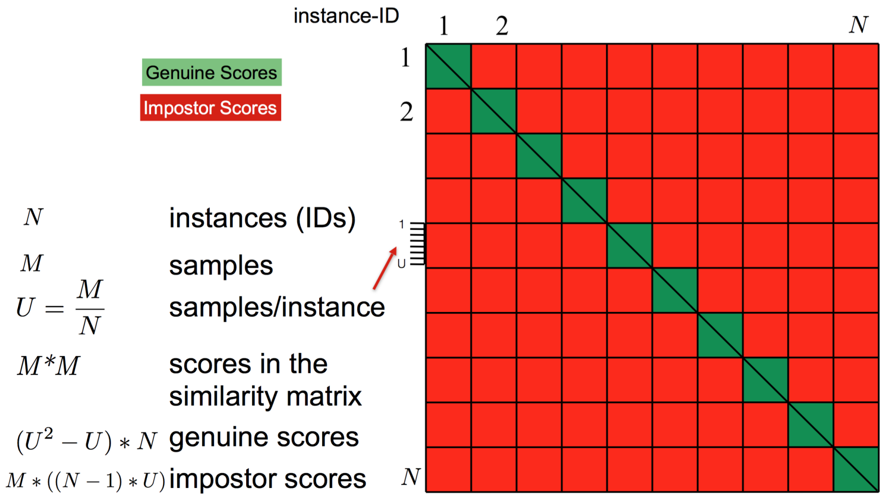

instance-ID
1
2
N
1
2
N
🟩 Genuine Scores
🟥 Impostor Scores
1
U
N instances (IDs)
M samples
$U = \frac{M}{N}$ samples/instance
$M*M$ scores in the similarity matrix
$(U^2 - U) * N$ genuine scores
$M * ((N-1) * U)$ impostor scores

Figure 2.6: Similarity matrix for 3 subjects and 3 instances

- Total number of instances N: $1.500 = S * 10$
- Number of samples per instance U: 12
- Total number of samples M: $18.000 = 12 * 1.500 = U * N$
- Number of genuine scores per instance: $132 = 12^2 - 12 = U^2 - U$
- Total number of genuine scores: $198.000 = (12^2 - 12) * 1.500 = (U^2 - U) * N$
- Total number of impostor scores: $323.982.000 = 18.000*149*12 = M*(N-1)*U))$

As we realize the total number of impostor scores is by two magnitudes larger than the number of genuine scores. Thus in order to avoid bias towards impostor scores it is common practice to reduce the impostor scores to corresponding instances from different subjects. For example one would include in the non-mated comparison trials only those samples, that are captured from the equivalent finger types (e.g. from the left index finger).

On the way to progress from comparison scores to performance measures we define for a given corpus of biometric samples as follows:

- Ω_g: set of all genuine scores
- Ω_i: set of all impostor scores
- $\Omega_g(t)$: set of all genuine scores $> t$
- $\Omega_i(t)$: set of all impostor scores $> t$
- $|| \, \Omega \, ||$: number of elements in Ω

Then we can compute

$$FMR(t) = \frac{\Omega_i(t)}{\Omega_i} \tag{2.17}$$

$$GMR(t) = \frac{\Omega_g(t)}{\Omega_g} \tag{2.18}$$

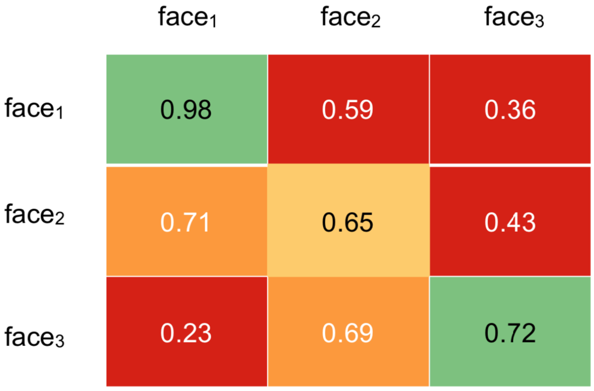
and then

$$FNMR(t) = 1 - GMR(t) \tag{2.19}$$

2.4.4 Example

Now we apply the measures from the previous section on the example, which was introduced in Figure2.5.

First in Figure 2.8 we set the threshold t to 0,66.



	face₁	face₂	face₃
face₁	0.98	0.59	0.36
face₂	0.71	0.65	0.43
face₃	0.23	0.69	0.72

Figure 2.7: Performance metrics for threshold t=0,66

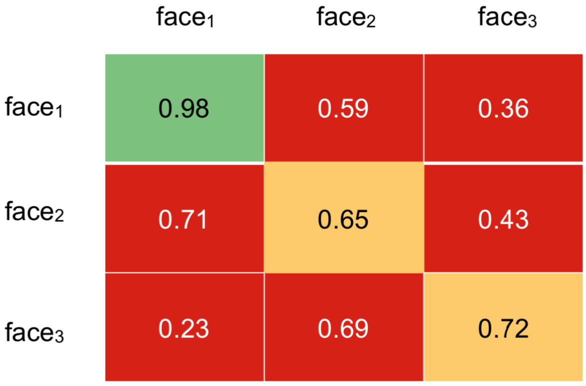
We can observe that 2 impostor scores exceed the threshold. Thus we compute according to equation 2.17 the $FMR(0,66) = 2/6$ and according to equation 2.18 the $FNMR(0,66) = 1/3 = 1 - 2/3$

Next we modify the threshold and evaluate the system for $t = 0,73$.

Now we can observe that no impostor scores is exceeding the threshold. But as a consequence of the modified threshold now tow genuine comparison trials are not successful any longer. We compute the $FMR(0,66) = 0$ and the $FNMR(0,66) = 2/3 = 1 - 1/3$. We can clearly identify the dependency between FMR and FNMR.

2.5 Reporting

In order to benchmark various algorithms we use a graphical representation plotting the GMR/FNRM in relationship to the FMR for different thresholds t. In Figure 2.9 we



	face$_1$	face$_2$	face$_3$
face$_1$	0.98	0.59	0.36
face$_2$	0.71	0.65	0.43
face$_3$	0.23	0.69	0.72

Figure 2.8: Performance metrics for threshold t=0,73

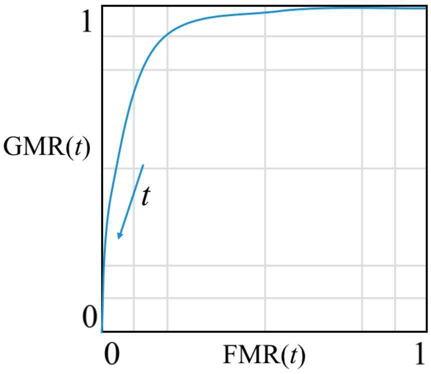


Figure 2.9: Receiver Operating Characteristic (ROC)

observe the Receiver Operating Characteristic (ROC) and most recommended graphical reporting in Figure 2.10 the Detection Error Trade-Off (DET) Curve.

The benchmark of various algorithms (biometric comparison subsystems) is illustrated in Figure 2.11. We can say that technically speaking that system performs best that shows the lowest area under curve. Note that this is not System D.

2.6 Verification System Performance

The first order estimation of the performance for a verification system that is based on transactions allowing multiple attempts can be derived from the detection error trade-off curve. However if this is applied the potential correlations between the attempts are neglected. Such correlations could be due to habituation of the capture subject with the human- computer interface of the biometric system. The relevant measures for a verification system are the False-Accept-Rate (FAR) and the False-Reject-Rate (FRR). The ISO-definition [7] for both metrics are the following:

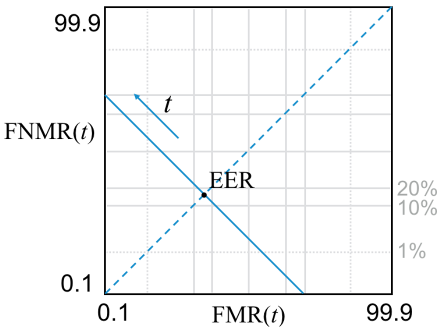


Figure 2.10: Detection Error Trade-Off (DET) Curve

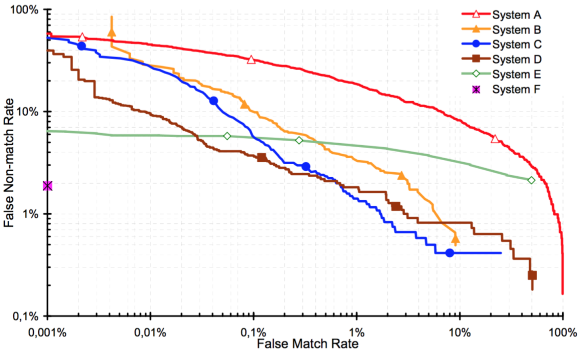


Figure 2.11: Detection Error Trade-Off (DET) Curve

False-Accept-Rate (ISO/IEC 19795-1): *proportion of verification transactions with wrongful claims of identity that are incorrectly confirmed.*

False-Reject-Rate (ISO/IEC 19795-1): *proportion of verification transactions with truthful claims of identity that are incorrectly denied.*

For the simplified case that the verification system does allow only a single attempt per transaction then the FAR and FRR can be estimated as follows.

$$FAR = FMR * (1 - FTA) \tag{2.20}$$

and

$$FRR = FTA + FNMR * (1 - FTA) \tag{2.21}$$

If the biometric application is likely to be confronted with a large number of failure to enrol cases (e.g. as it is a fingerprint system for mine workers) and the biometric performance shall be predicted based on a gallery that was collected for a technology testing then the equations 2.20 and 2.21 do not sufficiently express the performance to be expected. The reason for this is that in a technology evaluation biometric references are generated from the gallery that do not cause a failure-to-enrol and probes that do not cause a failure-to-acquire. For such a case the generalized versions of the above equations are more appropriate, which are given by:

$$GFAR = FMR * (1 - FTA) * (1 - FTE) \tag{2.22}$$

and

$$GFRR = FTE + (1 - FTE) * FTA + (1 - FTE) * (1 - FTA) * FNMR \tag{2.23}$$

2.7 Identification System Performance

The first order estimation of the false positive and false negative identification rates for open-set systems, can be derived from FMR and FNMR and the DET curve. However, such estimates cannot take account of correlations in the comparisons involving the same data subject, and consequently can be quite inaccurate [7].

$$FPIR = (1 - FTA) * (1 - (1 - FMR)^N) \tag{2.24}$$

where $FPIR$ is the False-Positive-Identification-Rate. For a small FMR we can substitute in equation 2.24

$$(1 - FMR)^N \approx 1 - N * FMR \tag{2.25}$$

and thus under the assumption of $FTA = 0$ we derive

$$FPIR = (1 - 0) * (1 - (1 - N * FMR) \tag{2.26}$$

$$FPIR = N * FMR \tag{2.27}$$

This is comparable to the scenario that has been conducted at the Mainz as well as Berlin Suedkreuz trainstation

2.8 Testing Standards

Test procedures as such are well known since the biometric performance testing standards ISO/IEC 19795-1 was established in 2006 [3] and then revised in 2021 [7]. That framework for Biometric Performance Testing and Reporting was developed on the basis of established concepts such as the *Best Practices in Testing and Reporting Performance of Biometric Devices*[11] and it defines in which way algorithm errors such as false-match-rate (FMR) and false-non-match-rate (FNMR) as well as system errors such as false-accept-rate (FAR) and false-reject-rates (FRR) must be reported.

2.9 Reading

Complementary reading for performance testing is the ISO standard on Biometric Performance Testing [7]. Details on the rule of 3 are given in the paper of Jovanovic and Levy [10].

Bibliography

[1] ISO/IEC JTC1 SC37 Biometrics. *ISO/IEC 19794-4:2005 Information Technology – Biometric Data Interchange Formats – Part 4: Finger Image Data.* International Organization for Standardization, July 2005.

[2] ISO/IEC JTC1 SC37 Biometrics. *ISO/IEC 19794-5:2005. Information Technology - Biometric Data Interchange Formats - Part 5: Face Image Data.* International Organization for Standardization, June 2005.

[3] ISO/IEC JTC1 SC37 Biometrics. *ISO/IEC 19795-1:2006. Information Technology – Biometric Performance Testing and Reporting – Part 1: Principles and Framework.* International Organization for Standardization, March 2006.

[4] ISO/IEC JTC1 SC37 Biometrics. *ISO/IEC SC37 SD11 General Biometric System.* International Organization for Standardization, May 2008.

[5] ISO/IEC JTC1 SC37 Biometrics. *ISO/IEC 30107-1. Information Technology - Biometric presentation attack detection - Part 1: Framework.* International Organization for Standardization, 2016.

[6] ISO/IEC JTC1 SC37 Biometrics. *ISO/IEC 2382-37:2017 Information Technology - Vocabulary - Part 37: Biometrics.* International Organization for Standardization, 2017.

[7] ISO/IEC JTC1 SC37 Biometrics. *ISO/IEC 19795-1:2021. Information Technology – Biometric Performance Testing and Reporting – Part 1: Principles and Framework.* International Organization for Standardization, June 2021.

[8] Jain, A., Flynn, P., and Ross, A. *Handbook of Biometrics.* Springer, July 2007.

[9] Jain, A., Ross, A., and Pankanti, S. Biometrics: A tool for information security. *IEEE Trans. on Information Forensics and Security* (2006).

[10] Jovanovic, B., and Levy, P. A look at the rule of three. *The American Statistican* (1997), 137–139.

[11] Mansfield, T., and Wayman, J. Best practices in testing and reporting performance of biometric devices. CMSC 14/02 Version 2.01, NPL, August 2002.