IMTC 2004 – Instrumentation and Measurement
Technology Conference
Como, Italy, 18-20 May 2004

# Accuracy and Performance of Biometric Systems

Marco Gamassi, Massimo Lazzaroni, Mauro Misino, Vincenzo Piuri, Daniele Sana, Fabio Scotti
Department of Information Technologies, University of Milan
Via Bramante 65, 26013 Crema (CR), Italy
Phone: +39-02-503-30066, Fax: +39-02-503-30010, Email: {gamassi, lazzaroni, misino, piuri, sana, scotti}@dti.unimi.it

***Abstract*** *– Although efforts of the entire international biometric community, the measurement of accuracy of a biometric system is far to be completely investigated and, eventually, standardized. The paper presents a critical analysis of the measurement of accuracy and performance of a biometric system. Current approaches to the problem and procedural error have been described and criticized. Finally, a methodology for the measurement of the accuracy of biometric systems with not-symmetric matching function is proposed and discussed.*

***Keywords*** *– Biometric Systems, fingerprint, iris recognition, best practice, accuracy measurement.*

## I. INTRODUCTION

Biometric systems have been defined by the USA National Institute of Standards and Technology [1] as systems exploiting "automated methods of recognizing a person based on physiological or behavioral characteristics" (*biometric identifiers*, also called *features*). Physiological biometrics is based on data derived from direct measurement of a body part (i.e. fingerprints, face, retina, iris), while behavioral biometrics is based on measurements and data derived from a human action [2] (i.e. gait and signature).

Biometric systems are being used to verify identities and restrict access to buildings, computer networks, and other secure sites [3]. Recent global terrorism escalation is pushing the need for secure, fast and non-intrusive identification of people as a primary goal for homeland security. As commonly accepted, biometrics seems to be the first candidate to efficiently satisfy these needs. For example, by October 2004 USA planned to control the accesses to/from country borders by means of biometric passports [4, 5].

Personal identification has taken the form of token-based or knowledge-based methods, such as secret passwords and PINs (Personal Identification Numbers), ID cards, keys, passes *etc.* Biometric approach completely differs from traditional methods since the identification is based on personal and unique peculiarities of individuals, which cannot be easily misplaced, forged, or shared [6].

Given that a biometric system is an identification system, its accuracy can be evaluated by classical techniques [7] but peculiarities are present. Typically, to effectively test biometric systems, a great number of volunteers is required or a large database of biometric records must be accessed [7, 8 , 9]. Experiments are complex, expensive and they expose the data maintainer to important problems related to the security and privacy of the biometric records. Furthermore, the protocol of the experiments can directly affect system accuracy [9, 10] and it is not possible to resume the overall system performance in a single index of accuracy to simply compare two different biometric systems.

This paper aims to present a critical analysis of the accuracy and performance measurement methodology of a biometric system and it proposes how to extend the measurement methodology in order to consider biometric systems that have a not-symmetric matching function. Section II presents the more frequently studied biometric systems in the literature and their peculiarities. Section III introduces the terms and the theory of the measure of accuracy of a biometric system. Section IV describes and criticizes current best practices as well as it proposes how to evaluate non-symmetric matching function systems into the comprehensive framework of accuracy evaluation. Finally, section V presents statistical considerations concerning the interval of confidence of the accuracy estimation and typical errors in setting up the biometrics experiments.

## II. BIOMETRIC SYSTEMS

From the literature a biometric system has a general structure. Figure 1 shows the components of a biometric system according to [10]. First of all, a sensor acquires a *sample* of the user presented to the biometric system (i.e. fingerprint, face, iris images). The sample can be transmitted, eventually exploiting compression/decompression techniques. Some systems store the complete sample data in the storage unit. Storing samples is often deprecated in the literature due to privacy and security issues [11, 12].

More correctly, a biometric system uses and stores only a mathematical representation of the information extracted from biometric samples by the signal processing module: the *biometric feature*. Examples are minutiae coordinates and iris-codes. If the extracted feature is stored (enrolled) into the biometric system, a *template* for future identification or verification (matching) is added. Each biometric system has a measure of the similarity between features derived from a presented input sample and a stored template. The measure produces the *matching score*. Hence, a match/non-match decision may be made according to whether this score exceeds a *decision threshold* or not. The term *transaction* refers to an user attempt to validate a claim of identity (or non-identity) by consecutively submitting one or more samples, as allowed by the system decision policy [10].

The signal processing module represents the core of the system and is generally composed by sub-modules which implement the preprocessing functions (i.e. image filtering and enhancement), the feature extraction and the matching between two features.

Typically a biometric system can be characterized by the following attributes: uniqueness, universality, permanence, measurability, user friendliness, acceptability and circumvention [10]. *Uniqueness* refers to the fact that a feature must be unique: an identical feature should not appear in two different people. *Universality* means that the feature type is present/occurs in as many people as possible. Unfortunately we can not assume, for example, that every individual has all the fingers or has both irises not damaged. The *Permanence* property is related to the need that the feature does not change over time, or at least, it varies very slowly. *Measurability* concerns the possibility to measure the feature with relatively simple technical instruments. *User friendliness* requires that the measure should be easy and comfortable to be done, and *Acceptability* refers to the people's acceptance of the measure in daily lives. *Circumvention* concerns the toughness to deceive the system by fraudulent methods. All these attributes must be taken into account designing a biometric system.

Most cited biometric samples in the literature are: fingerprint, signature (hand-writing), facial geometry, iris, retina, hand geometry, vein structure, ear form, voice, DNA, odor (human scent), keyboard strokes and gait [2]. Each of them has different accuracy, cost and a different fulfillment of the seven attributes previously presented.

A biometric system can work basically in two configurations: identification and verification. *Identification* means that the acquired and processed biometric feature is compared to *all* biometric templates stored in a system. If there is a match, the identification is successful, and the corresponding user name or user ID is put in output. *Verification* means that the user enters her/his identity into the system (i.e. by keyboard or using a card) and a biometric feature is scanned. Then, the system compares the input feature *only* with the previously enrolled reference feature corresponding to the ID. If a match occurs, verification is successful. Systems that use a single biometric feature are defined as *monomodal*. When the identification is computed by comparing the matching values between N biometric features different in type with a specific policy, the system is called *multimodal* [13]. Example of combinations such as face/fingerprint, iris/fingerprint, and face/voice are particularly discussed in the literature [13-15]. Many studies report an improvement in accuracy for multimodal systems with respect to systems working with single biometric features [14-16].

## III. BIOMETRIC SYSTEM EVALUTATION

The evaluation of a biometric system can be performed from different perspectives named: *technology*, *scenario* and *operational*. In this paper we deal with the technology evaluation since its goal is to compare competing algorithms when a sensor technology has been selected [7,10].

The *scenario evaluation* aims to determine the overall performance of a complete system in a prototype or simulated application that models a real-world target application. Since each tested system has its own acquisition sensor, it will receive slightly different data even if we acquire samples from the same individuals. Test results will be repeatable only if the simulated scenario can be carefully controlled. The *operational evaluation* tests a complete biometric system in a specific application environment with a specific target population. In general, operational test results will not be repeatable. The *technology evaluation* compares algorithms on a standardized database collected by a "universal" sensor. Of course, performance with this database will depend upon both the environment and the population in which it has been collected. Typically to avoid malicious approaches by the developers, it is possible firstly to provide them only a portion of the sample database, and distribute actual evaluation samples only after the developing of the algorithm's code. Testing is carried out using offline processing of the data. Because the database is fixed, the results of technology tests are repeatable.

Figure 2 shows the most general situation in a biometric database: we have a different number of samples for different individuals. Databases for algorithms comparison are poorly available [1, 17-20] due to the fact that they are very expensive and they contain complete biometric samples of real individuals. Security and privacy expects are seriously involved [11,12]. Some synthetic databases/generators are available only for fingerprints [21].

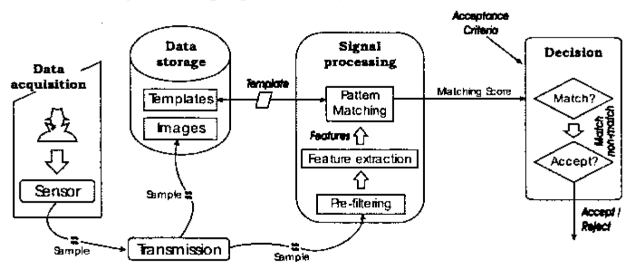



Figure 1: Structure of a biometric system.

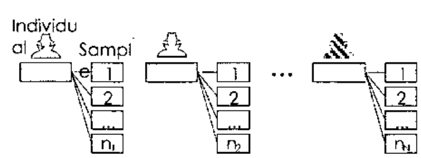



Figure 2: General samples situation of a biometric dataset

## IV. ACCURACY AND PERFORMANCE INDEXES

In the case of a *technology evaluation*, the accuracy indexes most commonly accepted in the literature are now discussed.

The following definition of accuracy presents differences with respect to the classical one used in metrology [22] but it is generally accepted in biometric systems. Accuracy of measurements evaluates the agreement between the result of a measurement and the expected value, applying the system on a standardized database, as described in the previous section.

In this paper, accuracy is given by indexes evaluated using the concept of error: this definition is typically used in biometric systems. Readers often confuse this measure of accuracy processed on a standard database with the accuracy of the methodology. However, at least a second source of uncertainty – which affects the overall accuracy – should be considered: the uncertainty introduced by the measurement process due, for example, to pressure, humidity, finger position, electronic noise, quantization, etc. [5]. The authors consider this second source of uncertainty of great interest and it will be the goal of the further research. Moreover, taking into account both methodological and measurement uncertainty is not a trivial task. If the extracted biometric feature comes from an ideal sensor obtained by an ideal collection procedure, the methodological uncertainty should be equal to zero. However, in presence of noise corrupted samples, the preferred method minimizes the effect of noise source on the accuracy.

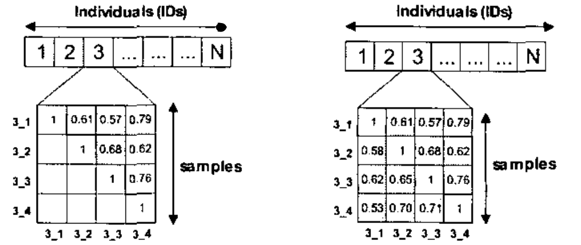
The following theory is valid for both monomodal and multimodal biometric systems. We can assume to have a sample database of identified individuals, as plotted in figure 2. In the literature many methods considered to evaluate the accuracy of a biometric system implicitly assume that the matching function is *symmetric* [15, 23 and 24]. Given two biometric features A and B and naming the matching function M, we have a symmetric matching function if M(A,B) = M(B,A). In the following we describe how to extend the equation for the accuracy evaluation for systems where we have M(A,B) ≠ M(B,A). Such systems are present in the literature, for example as described in [25] and [26]. In this paper, we do not comment if the symmetry is preferable to asymmetry in the matching function, but we will describe how is possible to make a fair comparison between different biometric systems by taking into account that issue.

Referring again to figure 2, let's define $B_{ij}$ as the $j^{th}$ sample of the $i^{th}$ individual (i.e. a fingerprint or iris image, either filtered or not); $T_{ij}$ as the template computed from $B_{ij}$ (the features extracted); $n_i$ as the number of samples available for the $i^{th}$ individual and, finally, N as the number of individuals enrolled. Let's follow the steps to compute the accuracy performance of the systems defining the proper indexes.

*A. Step 1 – Enrolment:*

The templates $T_{ij}$, where $i$=1..*N*, $j$=1..$n_i$, are computed from the corresponding sample $B_{ij}$ and stored on disk; if something wrong happens, an index (**$REJ_{ENROLL}$**) has to be increased.

$REJ_{ENROLL}$ is the rejection ratio in the enrolment phase, due to *Fail* (the algorithm declares it cannot enrol the biometric data), *Timeout* (the enrolment exceeds the maximum allowed time) and *Crash* (the algorithm crashes during biometric processing) situations [10,17].




a) *gms* for symmetric matchings b) *gms* for asymmetric matchings

Figure 3: Genuine Matching Scores.

| | 1_1 | 2_1 | 3_1 | ..._1 | N_1 |
|---|---|---|---|---|---|
| 1_1 | 1 | 0.11 | 0.27 | 0.19 | 0.09 |
| 2_1 | | 1 | 0.25 | 0.32 | 0.15 |
| 3_1 | | | 1 | 0.16 | 0.22 |
| ..._1 | | | | 1 | 0.20 |
| N_1 | | | | | 1 |

a) *ims* for symmetric matchings.

| | 1_1 | 2_1 | 3_1 | ..._1 | N_1 |
|---|---|---|---|---|---|
| 1_1 | 1 | 0.11 | 0.27 | 0.19 | 0.09 |
| 2_1 | 0.16 | 1 | 0.25 | 0.32 | 0.15 |
| 3_1 | 0.21 | 0.19 | 1 | 0.16 | 0.22 |
| ..._1 | 0.23 | 0.27 | 0.21 | 1 | 0.20 |
| N_1 | 0.14 | 0.08 | 0.26 | 0.28 | 1 |

b) *ims* for asymmetric matchings.

Figure 4: Impostor Matching Scores

*B. Step 2 – A general matching score computation:*

For *symmetric matching functions* the consuetude is as follows [17]: each biometric template $T_{ij}$ successfully created in the previous step is matched against the biometric sample $B_{ik}$ ($j<k\leq n_i$). The matching values are stored in a matrix called *Genuine Matching Scores* **$gms_{ijk}$** (figure 3.a). The term "genuine" refers to the fact that the matching is computed between samples of the same certified individual. Since the matrix is symmetric by definition, *only* the upper triangular matrix is computed. Each individual has its own squared gms matrix.

We now propose how to include systems that have asymmetric matching-function into the framework proposed in the literature. Next section considers statistic effects on the estimation of the systems accuracy.

For *asymmetric matchings* each biometric template $T_{ij}$ successfully created in the previous step is matched against the biometric sample $B_{ik}$ ($1\leq k\leq n_i$, $k\neq j$) and the corresponding *Genuine Matching Scores* matrix **$gms_{ijk}$** is stored (figure 3.b). The matrix is not symmetric but it is still square. Then, the number of matches, denoted as **NGRA** (*Number of Genuine Recognition Attempts*) is given by

$$NGRA_{symMatch} = \frac{1}{2}\sum_{i=1}^{N} n_i(n_i - 1) \quad (1)$$

where $REJ_{ENROLL}$ = 0 (symmetric matching)

$$NGRA_{asymMatch} = \sum_{i=1}^{N} n_i(n_i - 1) \quad (2)$$

where $REJ_{ENROLL}$ = 0 (asymmetric matching).

Let's now consider the matching values of samples of *different* individuals (*impostors matching*). For symmetric matching, each biometric template $T_{i1}$, $i=1..N$ is matched against the first biometric sample from different individual $B_{k1}$ ($i<k\leq N$) and then the corresponding *Impostor Matching Scores* $\mathbf{ims_{ik}}$ matrix is stored (Figure 4.a). Impostor matching in the case of asymmetric matching function is computed as follows: each biometric template $T_{i1}$, $i=1..N$ is matched against the first biometric sample from different individual $B_{k1}$ ($1\leq k\leq N$, $k \neq i$) and the corresponding *Impostor Matching Scores* $\mathbf{ims_{ik}}$ matrix is stored (Figure 4.b). The number of matches, denoted as **NIRA** (*Number of Impostor Recognition Attempts*) is given by

$$NIRA_{symMatch} = \frac{1}{2}N(N-1) \quad (3)$$

if $REJ_{ENROLL} = 0$ (symmetric matching), and

$$NIRA_{asymMatch} = N(N-1) \quad (4)$$

if $REJ_{ENROLL} = 0$ (asymmetric matching). Higher scores of matching values are associated with more closely matching images.

Finally, in the determination of *gms* and *ims* matrixes it is possible to have Fail, Timeout or Crash rejections. These events are respectively accumulated into $REJ_{NGRA}$ and $REJ_{NIRA}$ counters. It leads that *gms* and *ims* matrixes can have missing values. Commonly, in this case, special values are stored, i.e. "NULL" or negative matching values.

### *C. Step 3 – Accuracy Indexes*

In this section we describe how to evaluate the confidence of the accuracy indexes, as defined in the literature, for a biometric system. Considering systems allowing multiple attempts or having multiple templates, a general definition defines errors of the matching algorithms considering *single* comparisons of a submitted sample against a *single* enrolled template. The rates are: False Match Rate **FMR(*t*)** and False Non-Match rate **FNMR(*t*)**. They are functions of the threshold value ***t*** used to compare the matching value to make the decision.

The False Match Rate is the expected probability that a sample will be falsely declared to match a single randomly-selected template (*false positive*). The False Non-Match Rate is the expected probability that a sample will be falsely declared not to match a template of the same measure from the same user supplying the sample (*false negative*) [9].

The **FMR(*t*)** and **FNMR(*t*)** curves are computed from *gms* and *ims* distributions for *t* typically ranging from 0 to 1. Given a threshold *t*, **FMR(*t*)** and **FNMR(*t*)** are defined as follows [16]:

$$FMR(t) = \frac{card\{ims_{ik} | ims_{ik} \geq t\}}{NIRA} \quad (5)$$

$$FNMR(t) = \frac{card\{gms_{ijk} | gms_{ijk} < t\} + REJ_{NGRA}}{NGRA} \quad (6)$$

where *card* represents the cardinality.

The evaluation of the overall accuracy level of a biometric system is often evaluated by considering two error plots. The first is the Receiving Operating Curve (**ROC**), where (1− **FNMR**) is plotted as a function of **FMR** for all available values of *t*. The second, and most used, is the plot of FNMR *vs.* FMR in a logarithmic chart, called the Detection Error Trade-off (**DET**) plot. Figure 5 shows patterns of the DET curves computed for 6 different systems [17]. The best system is the one that has its DET curve below all the others. It would mean that, for all the values of its threshold *t*, the system yields the lowest FMR and FNMR with respect to the others. Typically a system outperforms all the others in *some* intervals of threshold *t*, not for all the values. DET plots are suitable to compare biometric systems.

In order to evaluate the peculiar behaviour of a selected system in separating the genuine from the impostor attempts, the *distributions* of the matching function values of the genuine population ($\mathbf{gms_{ijk}}$) and of the impostor population ($\mathbf{ims_{ik}}$) can be plotted. The smaller the overlap (the darker area in Figure 6), the better the biometric system will be. If no overlap occurs, it means that it exists a threshold value ***t'*** which perfectly separates the genuine individuals from the impostors (ideal case).

Other error-indexes can complete the accuracy description. The **EER** (Equal Error Rate) is often considered, and it is computed as the point where **FMR(*t*)** = **FNMR(*t*)**. Score distributions are typically not continuous and the **EER** must be often interpolated by the quantized data [17].

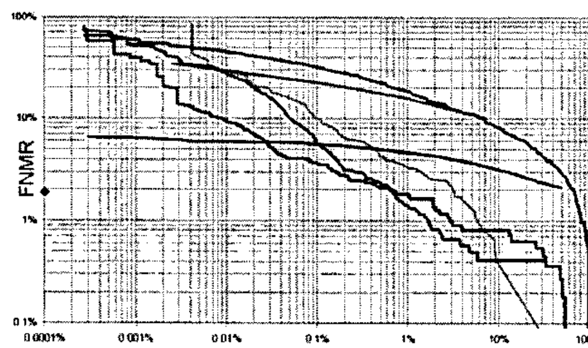



Figure 5: Examples of DET curve.

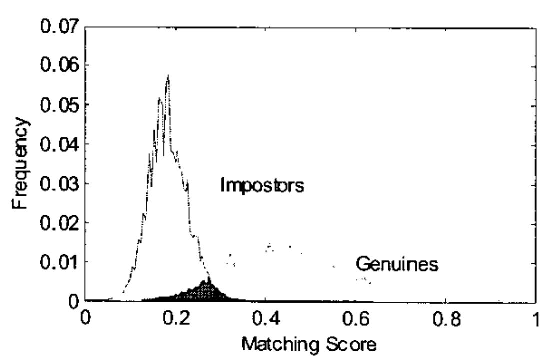



Figure 6: Examples of genuine and impostor distributions.

Other indexes measure the capability of the biometric system to *acquire* sample or to *process and enrol* templates: performance indexes. The former is the Failure to Acquire Rate (**FTA**) and it is "*the expected proportion of transactions for which the system is unable to capture or locate an image or signal of sufficient quality*" [10]. The latter is named Failure to Enrol Rate (**FER**) and it represents the "*expected proportion of the population for whom the system is unable to generate repeatable templates*" [10]. Examples are: individuals that are unable to present the required biometric feature, samples that have insufficient quality at enrolment, and those who cannot reliably match their template. For example, it has been estimated that about 2%–3.5% of individuals have their fingerprint ridges damaged by friction during a two-year period [20].

In order to shorter the matching time, some systems can sort/organize templates into bins. The Penetration Rate (**PR**) is defined as "*the expected proportion of the templates to be searched over all input samples under the rule that the search proceeds through the entire partition regardless of whether a match is found*" [10]. Of course, if the system fails to recognize the proper partition of a new sample we have a binning error. This proportion of misplaced samples represents the Binning Error Rate (**BER**).

In the literature many other indexes are present for testing biometric system's performances, but unfortunately they depend on the envisioned system's structure (identification/verification, fixed threshold, number of enrolled users and number of templates per user) [10]. This issue must be carefully taken into account comparing different systems [9]. The most common are **False Accept Rate** (FAR) and **False Reject Rate** (FRR). Considering also the Binning Error Rate (BER) and penetration rate (PR), and if the acceptance depends on a single successful match, we can write

$$FAR = PR \times FMR \times (1 - FTA) \quad (7)$$

$$FRR = FTA + (1 - FTA) \times BER + (1 - FTA) \times (1 - BER) \times FNMR \quad (8)$$

It is worth noting that it is a non-sense to describe the system performance by only its FAR or FRR. The two indexes must be both provided since they depend on the fixed threshold $t$: changing $t$ it is possible to arbitrarily reduce one of the two.

## V. CONFIDENCE OF ACCURACY ESTIMATION

The evaluation of confidence of the accuracy computed in previous sections and its relationship to the dataset size are now discussed. The proposed approach and definitions are generally used when describing a biometric system (see for example ref. 9). In general, "a N% *confidence interval* for parameter $x$ consists of a lower estimate $L$, and an upper estimate $U$, such that the probability of the true value being within the interval estimated is the stated value (e.g.: $P(x \in [L, U]) = N\%$)" [10]. Of course, the smaller the evaluation test size, the wider the confidence interval will be.

The "size" of an evaluation test can be thought in terms of the number of volunteers involved in the testing phase and the number of attempts made. The criterion used to choose volunteers/samples will influence how accurately error rates can be measured. In the literature, the term "*Non-self*" is used in the sense of "genetically different". It has been noted [27-29] that comparison of genetically identical biometric characteristics (for instance, between a person's left and right eyes or across identical twins) yields, on average, more similar score distributions than comparison of genetically different characteristics. Consequently, such genetically similar comparisons could not be considered in computing the false match rate.

It must be also noticed that the assumption about independency of all trials is not always satisfied (i.e. asymmetric/symmetric matching values in the ***igm*** matrix, problem related to "*Non-Self*" samples). The alternative is to compromise the independence of the samples by reusing a subset of all the volunteers and to expect a loss of statistical significance [10]. The actual consequence of not-independent samples in the test-database for a biometric system is not well understood yet [9].

Furthermore, performance estimates will be affected by both systematic errors and random errors. In biometric systems, by definition, random errors are due to the natural variation in people employed in the test, samples *etc.* Instead, systematic errors are due to bias in the test procedures, *etc.* For example, if certain types of individuals are under-represented in the volunteer set, this can give rise to a "bias" in the results [10]. It is fundamental to reduce the bias as much as possible and to report it into the results of the analysis. This allow for further fair comparisons between experiments.

It is interesting to note that some biometric producers state part-per-million (p.p.m.) errors in their systems, but errors in the data-collection procedures are typically considered much higher (due to "human errors" or factors such as iris/fingertips illness/injures previously described) [9, 20].

Dimensioning the test size, two main rules can be followed. They are known in the literature as the *rule of 3*, and the *rule of 30*. The *Rule of 3* [30-32] addresses the question "What is the lowest error rate that can be statistically established with a given number N of independent comparisons?". This value is the error rate $p$ for which the probability of zero errors in N trials is, for example, 5%. This gives $p \approx 3/N$, for a 95% confidence level. For example, a test of 300 independent samples returning no errors can be said with 95% confidence to have an error rate $\leq$ 1% [10]. The *Rule of 30* [33] is utilized to determine the evaluation test size and it can be expressed as follows: "To be 90% confident that the true error rate is within ±30% of the observed error rate, there must be at least 30 errors". So, for example, if we have 30 false non-match errors in 3,000 independent genuine trials, we can say with 90% confidence that the true error rate is between 0.7% and 1.3%. These rules have been derived from the *binomial distribution* assuming independent trials, and may be applied by considering the performance expectations for the evaluation. The two rules should be considered as over-optimistic [9].

Using a number of samples sufficiently large, the *central limit theorem* [34] implies that the observed error rates should follow an approximately *Gaussian (or normal) distribution.* Under the assumption of normality, 100 • (1-α) % confidence bounds on the observed error rates are given by the following formula:

$$\hat{p} \pm z \cdot \left(1 - \frac{\alpha}{2}\right) \cdot \sqrt{\hat{V}(\hat{p})} \qquad (9)$$

where:

$\hat{p}$ is the observed error rate and

$\hat{V}(\hat{p})$ is the estimated variance of observed error rate [9],

$z(\,)$ is the inverse of the standard normal cumulative distribution.

For 95% confidence limits the value z(0.975) is 1.96.

Often this formula gives rise to negative values for the error rate – but negative error rates are impossible. This is due to non-normality of the distribution of observed error rates. When a case like that occurs, non-parametric methods, such as the *bootstrap* [35], can be used to obtain confidence intervals.

Finally, it must be noted that a biometric system is not more accurate just because it uses a more complicated feature than other systems. Statements such as "iris biometrics is more accurate than fingerprint because its biometric feature is much more complicated" are not correct. Under quite general assumptions, in [37] has been demonstrated that the accuracy does not depend only on the number of degrees of freedom of the biometric features utilized. Of course, the accuracy depends on *how* information of the biometric features is used, much more than the "complexity" of the biometric features.

## VI. CONCLUSIONS

In this paper we summarize and critically discuss the main issues to be taken into account for the evaluation of the accuracy and performance of a biometric system. The case of a technology evaluation has been considered according to current best practices. The discussed methodology has a general appliance to different samples database formats and we propose how to support asymmetric matching algorithms. Our analysis enlightens that more efforts should be done to analyze the accuracy of the biometric systems from a stricter metrological point of view. The estimation of uncertainty in biological and clinical measurements is a true critical point and will be considered with a deeper metrological approach.

## REFERENCES


[1] http://www.nist.gov/srd/biomet.htm

[2] The Biometric Consortium (2002) 'Introduction to Biometrics', http://www.biometrics.org/html/introduction.html

[3] J.D.M. Ashbourn, 'Biometrics : Advanced Identify Verification: The Complete Guide', Springer-Verlag, 2000

[4] S. Waterman, 'Biometric borders coming', The Washington Times, 27 Oct. 2003.

[5] United States General Accounting Office, 'Technology Assessment: Using Biometrics for Border Security', GAO-03-174, November 2002,

[6] A.K. Jain, et al., "Biometrics: Promising frontiers for emerging identification market", computer , vol. 33, no. 2, February 2000.

[7] The Biometric Evaluation Methodology Working Group, 'Common Methodology for Information Technology Security Evaluation', 2002

[8] P. J. Phillips, A. Martin, W.M. Przybocki, "An introduction to evaluating biometric systems", IEEE computer , vol. 33, no. 2, February 2000

[9] V.S. Valencia, 'Biometric Testing: It's Not as Easy as You Think', Biometric Consortium Conference, Sept. 2003, Arlington VA USA.

[10] A. J. Mansfield, J. L. Wayman, "Best practices in testing and reporting performance of biometric devices", Version 2.01, Aug. 2002.

[11] R. Clarke, 'Biometrics and Privacy', Proc. Computers, Freedom & Privacy 2002, San Francisco, April 2002.

[12] M. Crompton, 'Biometrics and Privacy', Proc. 1st Biometrics Institute Conference, 20 March 2003.

[13] R. Frischholz, U. Dieckmann, "BioID: A Multimodal Biometric Identification System", in IEEE Computer, Vol. 33, No. 2, pp. 64-68, Feb. 2000

[14] A. Ross and A. K. Jain, "Information Fusion in Biometrics", Pattern Recognition Letters, Special Issue on Multimodal Biometrics, Vol. 24, No. 13, pp. 2115-2125, September 2003.

[15] A. Kumar, et al., "Personal Verification Using Palmprint and Hand Geometry Biometric", in Proceedings of Fourth International Conference on AVBPA, (Guildford, U. K.), pp. 668-678, June 2003

[16] D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar, Handbook of Fingerprint Recognition, Springer Verlag, June 2003

[17] D. Maio, D. Maltoni, R. Cappelli, J.L. Wayman and A.K. Jain, "FVC2002: Second Fingerprint Verification Competition", in proceedings 16th International Conference on Pattern Recognition (ICPR2002), Québec City, vol.3, pp.811-814, August 2002

[18] FPVC2000 database: http://bias.csr.unibo.it/fvc2000/databases.asp

[19] CASIA database: http://www.sinobiometrics.com/casiairis.htm

[20] R. Michael McCabe, 'Standards for Certifying Biometric Accuracy', BIOCONS2003.

[21] R. Cappelli, D. Maio and D. Maltoni, "Synthetic Fingerprint-Database Generation", Proc. of ICPR2002, Québec City, vol.3, pp.744-747, 2002.

[22] "International vocabulary of basic and general terms in metrology" published jointly by the International Bureau of Weights and Measures, the International Organization for Standardization, IEC, OIML, IFCC, IUPAC and IUPAP.

[23] X. Jiang; W.Y. Yau, 'Fingerprint minutiae matching based on the local and global structures', Proc. of ICPR2000,Vol. 2, 2000

[24] Jain, A.K., L. Hong, S. Pankanti, and R. Bolle, "An Identity Authentication System Using Fingerprints," Proceedings of the IEEE, Vol. 85, No. 9, Sept. 1997, pp. 1365-1388

[25] Y.P. huang, et.al.,'An efficient iris recognition system', Proc. International Conf. on Machine Learning and Cybernetics, Nov. 2002

[26] L. Yu, K.Q.Wang, C.F. Wang, D. Zhang, 'Iris verification based on Fractional Fourier Transform', Proc.First International Conference on Machine Learning and Cybernetics, November 2002.

[27] H.H. Newman, et.al., Twins, Chicago University Press, 1937.

[28] J. Daugman, and C. Downing, 'Epigenetic randomness, complexity, and singularity of human iris patterns', Proc. of the Royal Society Biological Sciences, 2001, 268, 1737-1740.

[29] A.K. Jain, S. Prabhakar, and S. Pankanti, "On the similarity of identical twin fingerprints", Pattern Recognition, 2002, 35(11), 2653-2663.

[30] T. A. Louis, "Confidence intervals for a binomial parameter after observing no successes", The American Statistician, 1981, 35(3), 154.

[31] B. D. Jovanovic, P. S. Levy, "A look at the rule of three", The American Statistician, 1997, 51(2), 137-139.

[32] J. L. Wayman, "Technical testing and evaluation of biometric identification devices", Biometrics: Personal identification in networked society, edited by A. K. Jain, et al., Kluwer, 2000, 345-368.

[33] G. R. Doddington, et.al. "The NIST speaker recognition evaluation: Overview methodology, systems, results, perspective", Speech Communication, 2000, 31(2-3), 225-254.

[34] G. W. Snedecor, W. G. Cochran, Statistical methods, Iowa State University Press, 1967 (Sixth edition).

[35] K. V. Diegert, "Estimating performance characteristics of biometric identifiers", Proceedings of Biometrics Consortium Conference, San Jose, CA, June 1996.

[36] J. L. Wayman, 'Degrees of Freedom as Related to Biometric Device Performance', AVANTI: An International Journal of Biometrics, 1999.