Error-Rate Equations for the General Biometric System

Enhancing human-identification systems through multiple samples, multiple templates, and database-partitioning techniques.

by JAMES L. WAYMAN

The function of a biometric identification system is to verify claims of "customers" (users) that they are who they say they are, or are not who they say they are not. More specifically, the biometric system seeks to verify a customer's claim that her/his physiological and/or behavioral characteristics do or do not match those of some number of previously enrolled individuals. In the literature of biometric identification, a distinction is made between "verification" and "identification," "one-to-one" and "one-to-many" matching, based on whether the size of the searched database is one or more than one. Past testing [3-18] of biometric devices has focused on measuring "false acceptance" and "false rejection" rates, or developed "candidate lists" [14,15], in either "one-to-one" or "one-to-many" tests, often using unreported system decision policies. Device performance is often convolved with test design and system decision policy, making results difficult, or impossible, to compare between tests. What is needed is a consistent approach that clearly decouples device performance from the size of the search, the test design, and the decision policy.

The general biometric system allows a single user to enroll multiple measures or multiple presentations of the same measure and, during operation, to enter multiple samples for matching. Consequently, the general system may perform multiple comparisons, even when the customer is claiming to match a single identity. A single mathematical system model of throughput and error rates using common, testable measures can be constructed for the general "M-to-N" biometric system in which both the "one-to-one" and "one-to-many" models are seen as degenerate cases. Here, M refers to the number of samples submitted for each transaction, and N refers to the number of active templates or user models in the database. The M samples will be of one or more physiological or behavioral characteristics. Multiple characteristics might be acquired using different biometric technologies. We call this collection of samples a "sample ensemble."

There are U active, enrolled users and T stored templates or models for each. Like the submitted samples, the template ensemble will consist of one or more biometric characteristics. We call the set of T templates a "template ensemble." In this article, to limit complexity, we will consider the T templates in an ensemble to be independent. In reality, subtle interdependencies between the models may exist, such as when multiple fingerprints are used. The effect of the interdependence of the T models on the comparison error rate is very difficult to estimate from current data and, at the cost of inaccuracies in the equations developed, will largely be ignored in this article. We will leave for future studies the most general case of template ensembles containing multiple representations of several multiple, dependent characteristics. We will also not consider "cohort" modeling techniques [21], often used in speaker recognition systems, wherein single input samples are compared to multiple models in a closely related subset of users.

Although some systems allow the number of stored templates to vary over the enrolled individuals, in this article we will assume that T is fixed by system policy so that

$$N = T * U. \tag{1}$$

1070-9932/99/$10.00©1999IEEE

Depending upon the system enrollment policy, each of the templates might be created from a single enrollment sample, from multiple samples given in a single enrollment session, or as a weighted moving average of samples submitted during use over time.

Notation

N	number of active stored templates or models in the database
M	number of samples submitted during each transaction
m	number of samples used in an initial search
U	number of active, enrolled users
T	number of independent templates or models stored as an ensemble n
K	number of partitions in a filtering or binning method
B	number of binning and filtering methods
p_i	probability that a sample will be in the ith partition
P_i	penetration coefficient owing to the ith filter or binning method
P_{sys}	system penetration coefficient
ε_j	bin-error rate of the jth bin
ε_{sys}	system bin-error rate
D	similarity or distance measure (in device-dependent units)
$\Psi_G(D)$	"genuine" distance distribution function
$\Psi_I(D)$	"impostor" distance distribution function
$\Psi_T(D)$	inter-template distance distribution function
τ	similarity or distance score threshold (in device-dependent units)
FMR(τ)	false-match rate: the probability that a sample will be mistakenly matched with a nonself template.
FNMR(τ)	false-nonmatch rate: the probability that a sample will be mistakenly judged not to match a self-template when compared.
FNM$_i$	the probability that the ith sample will be falsely not matched because of binning or matching errors.
Q	number of matches required by decision policy to declare an identification
C	hardware-comparison rate (comparisons per unit time)
S	system-throughput rate (users per unit time)

In previous papers [1, 2, 19, 20], we developed a general system description and governing equations for the "one-to-one" and "one-to-many" systems. The goal of this article is to derive general error-rate and throughput equations for the more general "M-to-N" system under a variety of decision policies. It will be seen that error rates are strongly impacted by the system-throughput requirements; that is, that system speed and error rates are closely related.

Basic Measures

There are five important, inter-related measures that govern the performance of the general biometric system: 1) the "penetration coefficient," reflecting the expected portion of the enrolled ensembles to be compared to a single input sample; 2) the "bin-error rate," or probability that a search for a matching template in the database will be unsuccessful because the sample and template were erroneously placed in different "bins"; 3) the single-comparison false-match rate, or probability that an "impostor" template will be incorrectly matched to a sample; 4) the single-comparison false-nonmatch rate, or probability that a truly matching template will be missed; and 5) the comparison rate (sample-template comparisons per unit time) of the hardware, perhaps averaged over a time period long enough to include system availability considerations.

System Penetration Coefficient

Search efficiencies can be achieved by partitioning the N templates 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. During operation, submitted samples are compared only to templates in appropriate partitions, limiting the required number of sample-to-template comparisons. We refer to partitioning based on exogenous information as "filtering" and reserve the word "binning" for the use of endogenous information

Generally, a single template may be placed into multiple partitions if there is uncertainty regarding its classification. Some templates of extreme uncertainty as to classification are labeled as "unknown" and placed in all of the partitions. In operation, samples are classified according to the same system as the database, then matched against only those templates from the database that are in the same classification or classifications. The portion of the total database to be scanned, on average, for each search is called the "penetration coefficient" P, which can be defined as

$$P = \frac{E[\text{number of comparisons}]}{N} \tag{2}$$

where E[number of comparisons] is the expected number of comparisons required for a single input sample.

In estimating the penetration coefficient, it is assumed that the search does not stop when a "match" is encountered, but

continues through the entire partition. Of course, the smaller the penetration coefficient, the more efficient the system.

Methods used to partition the database will depend upon the operational purpose of the system. In "verification" systems, where the goal is to verify the customer's claim to a specified identity, template ensembles might be stored on a card in the customer's possession. For each transaction, the database is simply the T templates on the card. In other systems of similar purpose, the templates for all enrolled users are stored centrally. In such a system, it is possible to partition the database of N templates into U partitions, based on the claimed identity of the enrollee. In such a case, where templates are placed exclusively in one partition and each partition contains the same number of templates, the penetration coefficient P, owing to the filter is

$$P = \frac{1}{K}, \tag{3}$$

where K is the number of partitions. In the case of "verification" systems, where $K = U$, combining Eqs. (1), (2), and (3) shows that the expected number of comparisons required of an input sample is T, the size of the user's enrolled sample ensemble. The equations developed in this article, therefore, are independent of the architecture chosen for storage.

The equations developed in this article are independent of the architecture chosen for storage.

In "identification" systems where the goal is to verify the customer's claim to an unspecified enrolled identity, or the negative claim of no enrolled identity, we might consider the total database as T partitions of U templates each. Each template in each partition is linked to templates in each of the $T - 1$ other partitions through the identity of the enrolled user. For example, consider a system for verifying customers' negative claims of no enrolled identity in which fingerprint templates from left and right index fingers of each of U persons are stored. In this case, $T = K = 2$. Data in each partition will be linked by the identity of the enrollee. Separation of left and right prints is based on information not found in the prints themselves, so partitioning is a "filtering" operation performed at the time of enrollment. As in the previous example, the bins are exclusive and there is equality in the partition assignments, so Eq. (3) applies. By Eqs. (1), (2) and (3), the expected number of searches per input sample is seen to be U.

For more general filtering and binning, however, such as the partitioning of the database by gender, equality in partition size does generally not apply and Eq. (3) cannot be used. (Currently, only speaker verification can perform gender-based partitioning on the basis of information within the biometric pattern. For other technologies, gender-based filtering must be done on the basis of information given by the customer or by assessment of the system supervisory personnel.) A more general approach must be taken. If there are K partitions and p_i is the probability that a template is placed in the ith partition, then the ith partition will hold $N * p_i$ templates. If the samples and templates are from the same population, p_i is also the probability that the sample is in the ith partition. If a sample or template can only be placed in a single partition, then

$$\sum_{i=1}^{K} p_i = 1. \tag{4}$$

In cases where Eq. (4) holds and the partitions are exclusive (no "unknown" partition), the expected number of comparisons is

$$E[\text{number of comparisons}] = \sum_{i=1}^{K} p_i N p_i = N \sum_{i=1}^{K} p_i^2 \tag{5}$$

and the penetration coefficient P can be seen to be

$$P = \sum_{i=1}^{K} p_i^2. \tag{6}$$

We will now consider the case where the Kth bin represents an "unknown" classification. The unknown bin must always be searched and samples classified as "unknown" must be searched against all templates regardless of bin. Nonetheless, Eq. (4) continues to hold. So, the expected number of comparisons becomes

$$E[\text{number of comparisons}] = N * p_K + \sum_{i=1}^{K-1} p_i N(p_i + p_k) = N\left[p_K + \sum_{i=1}^{K-1}(p_i + p_K)p_i\right]. \tag{7}$$

The term in brackets on the right-hand side is the penetration coefficient under these conditions.

In the case of samples or templates of uncertain, but not completely unknown, classification, the general procedure is to place them into multiple bins, such that Eq. (4) does not hold. Rather,

$$\sum_{i=1}^{K} p_i > 1 \tag{8}$$

and Eqs. (5) through (7) do not hold. Calculation of the penetration coefficient as a function of bin probabilities p_i under the more general condition expressed by Eq. (8) has been given in [22]. Penetration coefficient can be calculated empirically from the binning assignments of both samples and templates by

$$P_{AVE} = \frac{\sum^{M}\text{samples}\sum^{N}\text{templates with bin(s) in common with sample}}{MN} \tag{9}$$

where P_{AVE} is the average, or expected value, taken over all users. Equation (9) is given in more mathematically precise, but less intuitive, form in [24].

There may be multiple, say B_i independent, filtering and binning methods used with each biometric measure in the ensemble. We will therefore add two subscripts to the penetration coefficient $P_{i,j}$ to indicate the ith measure and the jth binning or filtering method. If the methods are truly independent, the total penetration coefficient P_i for the ith measure, using B_i methods, can be written as

$$P_i = \prod_{j=1}^{B_i} P_{i,j}. \tag{10}$$

If correlations exist between any of the partitioning schemes, Eq. (10) will under-estimate the true penetration coefficient, meaning that the real penetration coefficient will be higher (worse).

The above equation applies to systems that use any ensemble size T. In those systems where $T > 1$ and $M = T$, partitioning can be done on the basis of the classification of the entire ensemble of the independent samples and templates. That is, a sample ensemble can be compared to only those template ensembles that are partitioned similarly on all measures. We call this "ensemble binning." The system penetration coefficient for the ensemble becomes

$$P_{\text{ensemble}} = \prod_{i=1}^{T} P_i. \tag{11}$$

Bin-Error Rate

The bin-error rate reflects the percentage of samples falsely not matched against the database because of inconsistencies in the binning process. This error rate can be easily measured by comparing binning partitions assigned for samples and matching templates. In general, the more bins that are used, the greater the probability that the bins will be inconsistently applied to truly matching measures. Errors are a function of the action of the bin classification algorithms on the input sample, and consequently, methodologies for inducing such errors are difficult to predict without a thorough knowledge of the algorithm.

Filtering errors, such as in the classification of an individual as "male" or "female," are due to mistakes in the externally collected data, generally made by human operators during the customer interview process. Like binning errors, filtering errors also cause samples to be falsely not matched to templates in the database. With filtering, however, system vendors can "externalize" these errors, blaming them on the data-collection process of the system administrator, not on the computational algorithms. Filtering allows the beneficial decrease in the penetration coefficient without the responsibility for the associated increase in false-nonmatch error rate. Individuals wishing not to be matched to previously enrolled templates can increase the probability of filtering errors through deliberate actions. Thus, the use of filtering can create system vulnerabilities to fraud that generally do not occur with binning. Because filtering errors are the result of inconsistencies in human judgement or deliberate fraud, they cannot be easily measured by purely technical tests and will not be considered in this article.

Binning errors can be measured by determining the percentage of truly matching biometric patterns that are placed by the system in noncommunicating bins. For each binning method employed, a single test can be designed to determine the binning penetration coefficient, by Eq. (9), and the bin-error rate ε calculated by,

$$\varepsilon = \frac{\text{number of inconsistently binned (sample} - \text{template) pairs}}{\text{number of (sample} - \text{template) pairs tested}}. \tag{12}$$

In a system using multiple binning methods on a single measure, not to make a bin-error requires that none of the individual binning methods produce an error. This awkward English actually best describes the underlying probabilistic relationship

$$1 - \varepsilon_i = \prod_{j=1}^{B_i} (1 - \varepsilon_{ij}) \tag{13}$$

where ε_i is the bin-error rate on the ith measure and ε_{ij} is the bin-error rate for the jth of the B_i binning methods used on that measure. Equation (13) assumes that bin-errors are independent. If the B_i binning methods for the ith measure have the same bin-error rate $\varepsilon_{i'}$, Eq. (13) can be rewritten as

$$\varepsilon_i = 1 - (1 - \varepsilon_{i'})^{B_i} = B_i * \varepsilon_{i'} - O(\varepsilon_{i'}^2) \tag{14}$$

where $O(\varepsilon_{i'}^2)$ indicates terms of order $\varepsilon_{i'}^2$ and smaller. For small $\varepsilon_{i'}$, as is the general case, Eq. (14) reduces to

$$\varepsilon_i \approx B_i * \varepsilon_{i'}. \tag{15}$$

For systems using ensemble binning, the ensemble penetration coefficient is calculated using Eq. (11) and the ensemble bin-error rate is calculated as

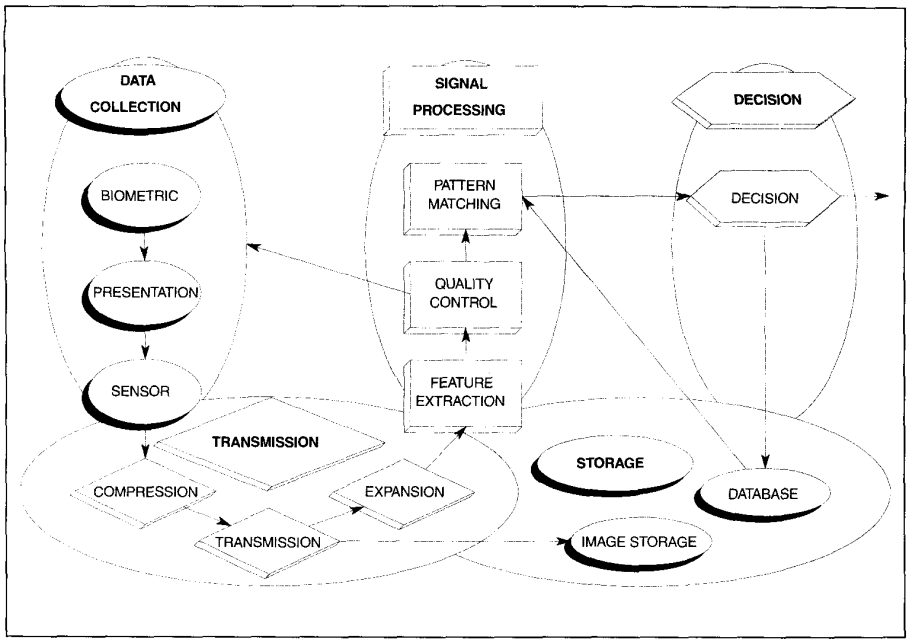

Figure 1. *The general biometric system.*

$$1-\varepsilon_{\text{ensemble}}=\prod_{i=1}^{T}(1-\varepsilon_i) \tag{16}$$

where the ε_i are the bin-error rates for the binning on each measure.

"Genuine," "Impostor," and "Inter-template" Distance Distributions

The function of the pattern-matching module in Fig. 1 is to send to the decision subsystem a positive, scalar measure D for every comparison of a sample to a template. We can presume, without loss of generality, that D increases with increasing difference between sample and template. We will loosely call this measure a "distance," recognizing that it will technically be such only if resulting from a vector comparison in a metric space. The general biometric system does not require that sample and template features compose such a space. (Minutiae-based fingerprint systems are an example of biometric system with sample and template features not composing a metric space. In general, fingerprint samples and templates will have unequal numbers of features, distances are not symmetric, and the triangular inequality does not hold.)

Regardless of the mathematical basis for the comparison, from a series of comparisons of samples to truly matching templates, we can construct a histogram that approximates the "genuine" distance probability distribution function [23]. We will call this distribution $\Psi_G(D)$, as shown in Fig. 2. It is both device- and measure-dependent. This "genuine" distribution is a measure of the repeatability of the biometric pattern. Repeatability is negatively impacted by any factor causing changes in the measurement. Such factors generally accumulate over time, so the "genuine" distribution appears to drift in the direction of increasing distance with the passage of time. This phenomenon is called "template aging," although changes in the sample, not the stored template, are responsible for this decrease in repeatability [24].

Similarly, from a series of comparisons of samples to different user's, or nonself, templates, we can construct a histogram that approximates the "impostor" distance probability distribution function, $\Psi_I(D)$. There are several different ways of doing this. The "impostor" histogram can be constructed by comparing each sample to a single nonself-template [12], by comparing every sample to every nonlike template [5,13-17], or through "resampling"[18, 25], which is drawing samples and templates from a pool at random with replacement. Some researchers [17] have suggested use of a "background" database of templates for which there is no matching sample. (In "large-scale" biometric systems with N exceeding 10 million, estimation of the false-match rate from $\Psi_I(D)$ becomes much more critical than estimation of the false-nonmatch rate from $\Psi_G(D)$. Further, single-comparison false-nonmatch rates of even 10% might give satisfactory system performance, while single-comparison false-match rates of 10^{-6} might be required. In this context, it is not possible to dismiss the need for a "background database" and the "independent degrees of statistical freedom" that ensue, provided that all templates and samples are collected from a similar population in a similar environment.

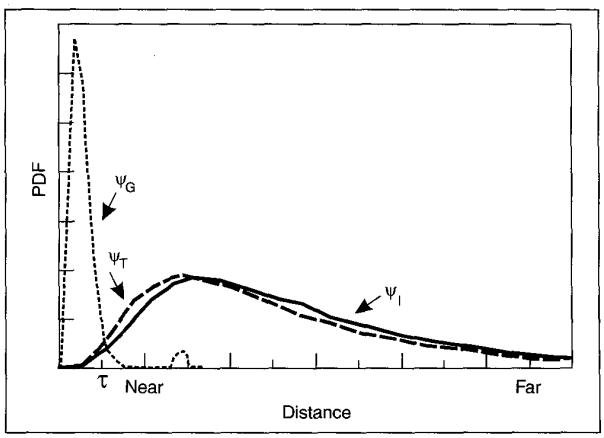

Figure 2. *Distance distribution functions.*

Consequently, for all approaches except that of comparing each sample to a single nonself template [12], the resulting histogram for the "impostor" distribution, by containing more measures, is much smoother than the histogram used to construct the "genuine" distribution, even if the same number of independent data points were used to construct both. However, the number of independent comparisons ("degrees of freedom") resulting from each method, needed for the development of confidence intervals, cannot exceed the number of independent samples. Calculation of confidence intervals is discussed in [24].

Ideally, the genuine and impostor distributions will be disjoint (nonoverlapping), allowing us to discriminate completely between "genuine" and "impostor" comparisons using a distance threshold. Of course, this is never the case in practice; one side of the problem being large distances between samples and truly matching templates caused by changes in the underlying biometric measure, in its presentation to the sensor, or in the sensor itself. We have noticed for a number of biometric devices that $\Psi_G(D)$ is usually bimodal, with the second mode coincident with the primary mode of $\Psi_I(D)$. This means that changes in the biometric measure, or its presentation, have caused a "genuine" individual to appear clearly as an impostor. We hypothesize that in the general system, the "genuine" distance does not increase smoothly with changes in the biometric pattern, but undergoes rapid increase with changes past a particular threshold. In any case, the general biometric system shows significant overlap in the tails of the "genuine" and "impostor" distributions.

There is actually a third distribution, the "inter-template" distribution $\Psi_T(D)$, that expresses the distinctiveness between the templates. In practice, only the templates and the (presumed) genuine comparison distances may be available to the researcher. The actual samples may be discarded by the system. In the case of a system that creates templates from but a single sample, templates are samples. In this case, the "inter-template" distribution is identical to the "impostor" distribution.

The general biometric system might use multiple samples taken at a single "enrollment" session to create the template or may update the template from a moving, weighted average of samples presented over time. Simulation models have shown us that, in these cases, the "inter-template" distribution is closer to the origin than the "impostor" distribution and, consequently, does not make a good proxy in calculating the relationship between "false match" and "false nonmatch" rates as a function of decision threshold. A mathematical development of the differences between the "impostor" and "inter-template" distributions is given in [24].

With the proper assumptions, we can construct the "impostor" distribution from the higher-dimensional convolution of the "genuine" and "inter-template" distributions. If we can assume that the "genuine" scalar distance measures result from an isotropic distribution of samples around the true templates, and that such distributions are "stationary," meaning that the distribution resulting from the set of single sample-to-template distances is the same as the distribution of each sample about its own template, then we can reconstruct the sample-to-template distance distribution from the genuine and inter-template distributions [26]. The reconstruction algorithm must account for the template creation policy. This is an area of current emphasis in our research and classifies as a "hard" problem.

The Single-Comparison False-Match Rate

A single-comparison false match occurs when a sample is incorrectly matched to a template in the database by the decision subsystem because the distance measure between the two is less than a threshold τ established by the decision policy. The single-comparison false-match rate $\mathrm{FMR}(\tau)$ can be computed from the integral of the "impostor" distribution function $\Psi_I(D)$ between zero and the threshold, as

$$\mathrm{FMR}(\tau) = \int_0^{\tau} \Psi_I(D)\, dD\,, \tag{17}$$

which increases with increasing decision threshold. Although in practice τ might be user dependent, our analysis will consider τ to be at a single, fixed value for all users. The single-comparison false-match rate can be seen in Fig. 2 as the area under Ψ_I between the origin and τ.

The Single-Comparison False-Nonmatch Rate

A single-comparison false nonmatch occurs when a sample is incorrectly not matched to a truly matching template by the decision subsystem because the distance between the two is greater than the fixed threshold. The single-comparison false-nonmatch rate $\mathrm{FNMR}(\tau)$ can be given as

$$\mathrm{FNMR}(\tau) = \int_{\tau}^{\infty} \Psi_G(D)\, dD = 1 - \int_0^{\tau} \Psi_G(D)\, dD \tag{18}$$

where $\Psi_G(D)$ is the genuine probability distribution function. $\mathrm{FNMR}(\tau)$ decreases with increasing decision threshold. The single-comparison false-nonmatch rate can be seen in Fig. 2 as the area under Ψ_G to the right of τ. It is clear from Eqs. (17) and (18) that false-match and false-nonmatch rates are competing factors based on the threshold. This trade-off can be determined based upon the comparative risks of system false-match and false-nonmatch errors.

Hardware-Comparison Rate

The "one-to-one," or "cold match," comparison rate C is the number of comparisons per second of a single sample to a single database template that can be made by the hardware. It is a function of the hardware-processing speed, the template size, and the efficiency of the matching algorithm. System availability must be considered when predicting the number of

comparisons that can be made over longer time periods, such as a day or a month.

Usually, the architecture for large-scale (large N) biometric systems is modular in the sense that processing speed can be designed to meet seemingly any requirement, although there are no doubt limits of scale as speed requirements get too great. In general, a single comparison may take as many as a few million operations. Measurement and prediction of system processing speed from component architecture or from direct measurement is discussed in [27] and will not be considered further in this article.

System Performance Equations

We are now in a position to write some first-principal equations reflecting the dependence of system performance on the parameters explained in the preceding section. By "system performance," we mean the timely and correct matching and nonmatching of customers to identities in a database of N template ensembles, based on a system decision policy utilizing M samples from each customer. In the development of the equations, we will assume that one sample of each independent measure is submitted, such that $M = T$. Departures from this assumption will be handled in the "Examples" section of this article. The possibilities for system decision policies are limited only by the imaginations of the system developers. We will develop equations capable of modeling the most common approaches. Owing to both complication and lack of data, we will ignore any and all correlations between errors. We will indicate, where possible, the impact of this simplification. Our goal will be to give system performance estimates and bounds, based on these simplifying, but admittedly inexact, assumptions.

In a multimeasure system with large N, search speed becomes an important issue. The usual approach is to conduct an initial search with a subset m of the collected samples M, where $m \leq M$ and $M = T$. This limited initial search will rule out most of the U enrolled users as potential matches with m, not M, comparisons for each, thus greatly increasing search efficiency.

Let's assume a system decision policy that requires, for a system "match" decision, Q matches of the M samples to a single enrolled ensemble of T templates. To do this, we will conduct an initial search of the relevant partitions of the database against m of the M samples. These m samples are searched sequentially through the entire database. Each of the initial m searches will result in $P_i * N$ comparisons and a total of $\sum_{i=1}^{m} P_i * N$ comparisons will be made over the m searches. Any matches found can be verified by comparisons of the remaining M samples against the remaining T templates from the same template ensemble. In other words, if the first of the m initial searches against the database produce no match, yet the second results in a match identifying a candidate template ensemble, the remaining $M-2$ input samples will be compared to the T templates in the identified ensemble. If this results in $Q-1$ or more matches, a system match is declared. Accordingly, if all m samples in the initial search falsely nonmatch, or more than $T-Q$ false nonmatches occur against a correctly matched enrolled ensemble, a system "nonmatch" is falsely declared.

We will allow each of the m samples in the initial search to have independent error rates, ε_i, $\mathrm{FMR}_i(\tau)$, and $\mathrm{FNMR}_i(\tau)$. This reflects the differences in the underlying "genuine" and "impostor" distributions for each sample and allows for sample-dependent, but not user-dependent, thresholds. For purposes of mathematical tractability, however, the samples in the remaining comparisons (which may include samples from among the m) will be assumed to have uniform error rates, $\mathrm{FMR}_{i'}(\tau)$ and $\mathrm{FNMR}_{i'}(\tau)$.

System False-Nonmatch Rate

When comparing a single input sample to a single stored template, for false nonmatch not to occur, there must be 1) no binning error and 2) no single-comparison false nonmatch. Assuming these errors to be independent, the probability of a correct match of a single sample with a truly matching template can be written as

$$\begin{aligned}\Pr[\text{correct match for } i\text{th sample}] &= 1 - \mathrm{FNM}_i \\ &= (1-\varepsilon_i)(1-\mathrm{FNMR}_i)\end{aligned} \tag{19}$$

where FNM_i is the probability that the ith sample will not be properly matched for any reason, and the explicit dependence of FNMR_i on threshold τ has been dropped for notational simplicity. Rewriting Eq. (19),

$$\mathrm{FNM}_i = \varepsilon_i + \mathrm{FNMR}_i - \varepsilon_i * \mathrm{FNMR}_i. \tag{20}$$

Some simple systems make a series of sample-to-template comparisons without using any ensemble concepts. The decision policy for such systems may only require a single match on one of these comparisons for a system match to be declared. A system false nonmatch occurs only when all m comparisons result in a false nonmatch. Assuming independence of false nonmatches,

$$\mathrm{FNM}_{\mathrm{sys}} = \prod_{i=1}^{m} \mathrm{FNM}_i. \tag{21}$$

The development of Eq. (21) assumes the comparisons to be independent. From elementary probability theory, we know that

$$\Pr[A \cap B \cap C \cdots] = \Pr[A] * \Pr[B \mid A] * \Pr[C \mid AB] \cdots \tag{22}$$

where $\Pr[B \mid A]$ indicates the conditional probability of event B occurring given that A has occurred. If $\Pr[A \mid B] = \Pr[A]$ and

$\Pr[C|AB] = \Pr[C]$, we say that events A, B, and C are independent. In practice, we find this not to be the case. In operational data, we have observed that a single-comparison false nonmatch increases slightly the probability that a subsequent biometric sample of the same characteristic from the same customer will also be falsely nonmatched. We can reasonably expect the probability of a false nonmatch to approach unity as the number of previous false nonmatches from the same session by the same customer increases. Further, we expect this hypothesis to hold for any reasonable threshold. Consequently, we expect that Eq. (21) will underestimate by some unknown amount the true system false-nonmatch rate.

For systems using ensembles of multiple measures, the ith of m searches against an entire ensemble to not result in a false nonmatch requires that: 1) the initial comparison of sample to template not result in a false nonmatch; and 2) $Q-1$ or more of the remaining patterns in the ensemble be correctly matched. Therefore, the probability of a correct identification being declared on the ith of the m sample comparisons is

$$\Pr[\text{correct ID declared on } i\text{th sample}] = (1-\text{FNM}_i)\sum_{j=Q-1}^{T-i}\binom{T-i}{j}(1-\text{FNM}_{U'})^j(\text{FNM}_{U'})^{T-i-j}. \tag{23}$$

The complement, that the correct identification is not declared on the as ith sample, can be given as

$$\Pr[\text{correct ID not declared on } i\text{th sample}] = 1-(1-\text{FNM}_i)\sum_{j=Q-1}^{T-i}\binom{T-i}{j}(1-\text{FNM}_{U'})^j(\text{FNM}_{U'})^{T-i-j}. \tag{24}$$

The concept expressed by Eq. (21) still applies, but with the more complicated definition of $\mathbf{FNM}_i$ given by Eq. (24). Assuming that the m searches are independent, the probability that a system false nonmatch occurs is, therefore

$$\text{FNM}_{\text{sys}} = \prod_{i=1}^{m}\left[1-(1-\text{FNM}_i)\sum_{j=Q-1}^{T-1}\binom{T-i}{j}(1-\text{FNM}_{U'})^j(\text{FNM}_{U'})^{T-i-j}\right] \tag{25}$$

where $\mathbf{FNMR}_{\text{sys}}$ is the system false-nonmatch rate.

For systems that use ensemble binning, the probability of a bin error is the same for all samples, namely $\varepsilon_{\text{ensemble}}$. If $\varepsilon_{\text{ensemble}}$ replaces ε_i in Eq. (19), then probabilities of correct match calculated by Eq. (19) will not be independent for each sample and cannot be used in developing the system false-nonmatch rate equation. When using ensemble binning, the bin error is not independent over the M comparisons, as each comparison looks in the same database partition. We can modify the above development by removing consideration of the binning error from Eq. (19), writing

$$\Pr[\text{correct match for } i\text{th sample}] = 1-\text{FMN}_i = 1-\text{FNMR}_i \tag{26}$$

so

$$\text{FNM}_i = \text{FNMR}_i. \tag{27}$$

Equations (21), (23), and (24) continue to hold using Eq. (27).

For the system to return a proper identification, we require: 1) no binning error for the entire ensemble and 2) no failure of all initial m searches to identify the ensemble. For a correct identification to be made, we write

$$1-\text{FNM}_{\text{sys}} = [1-\varepsilon_{\text{ensemble}}]\prod_{i=1}^{m}\left[1-(1-\text{FNM}_i)\sum_{j=Q-1}^{T-i}\binom{T-i}{j}(1-\text{FNM}_{U'})^j(\text{FNM}_{U'})^{T-i-j}\right] \tag{28}$$

Equation (28) can be rewritten as

$$\text{FNM}_{\text{sys}} = \varepsilon_{\text{ensemble}} + [1-\varepsilon_{\text{ensemble}}]\prod_{i=1}^{m}\left[1-(1-\text{FNM}_i)\sum_{j=Q-1}^{T-i}\binom{T-i}{j}(1-\text{FNM}_{U'})^j(\text{FNM}_{U'})^{T-i-j}\right]. \tag{29}$$

System False-Match Rate

In simple systems not using multiple independent measures arranged as ensembles, a match will be declared if any of the m sample-to-template comparisons over the entire database results in a match. Consequently, no system false match requires no single-comparison false match over the entire database:

$$1-\text{FMR}_{\text{sys}} = \prod_{i=1}^{m}[1-\text{FMR}_i]^{N*P_i}. \tag{30}$$

rewriting,

$$\mathrm{FMR}_{\mathrm{sys}} = 1 - \prod_{i=1}^{m} [1 - \mathrm{FMR}_i]^{N*P_i}. \tag{31}$$

For large $m * N * P$, the system false-match rate approaches 1 even for very small single-comparison false-match rates FMR_i. Consequently, such a system design cannot be used for large-scale "identification" systems.

In systems using ensembles of multiple measures, a system false match occurs if Q or more samples are falsely matched against the enrolled ensemble of a single individual. One general approach is to search $m \le M = T$ samples against a partition of the database. If any matches are found, the remaining samples are compared to the associated templates in the matched enrolled ensembles. If $Q - 1$ additional false matches are found in any single enrolled ensemble, a match will be falsely declared by the system. The probability that the ith of the initial m searches will result in a false match against a single nonmatching ensemble can be given by

$$\Pr[\text{false match on the } i\text{th sample}] = \mathrm{FNM}_i * \sum_{j=Q-1}^{T-i} \binom{T-i}{j} \mathrm{FNM}_U^{\,j} (1 - \mathrm{FNM}_U)^{T-i-j}. \tag{32}$$

Again the explicit dependence of the false-match rate on threshold τ has been dropped for notational simplicity and the false-match rates FMR_i within the summation sign are considered uniform.

Numerical computation of Eq. (32) from the single-comparison false-match rates for each sample FMR_i is straightforward, as the ensemble size, M, will always be small, perhaps reaching 10 in the case of a 10-print fingerprint identification system.

The search of the ith sample from the initial m patterns against the entire database will not result in a false match only if none of the $N * P_i$ searches end in a false match. Therefore, the probability that the ith of the m initial searches against the relevant partition of the database will not end in a false match is

$$\Pr[\text{incorrect ID not made on } i\text{th sample}] = \left[1 - \mathrm{FNM}_i * \sum_{j=Q-1}^{T-i} \binom{T-i}{j} \mathrm{FNM}_U^{\,j} (1 - \mathrm{FNM}_U)^{T-i-j}\right]^{N*P_i}. \tag{33}$$

For a search of a sample ensemble against the database to not end in a false match requires that none of the m initial comparisons falsely match. Therefore, the system false-match rate can be given as

$$1 - \mathrm{FMR}_{\mathrm{sys}} = \prod_{i=1}^{m} \left[1 - \mathrm{FMR}_i * \sum_{j=Q-1}^{T-i} \binom{T-i}{j} \mathrm{FMR}_U^{\,j} (1 - \mathrm{FMR}_U)^{T-i-j}\right]^{N*P_i} \tag{34}$$

which can be rewritten as

$$\mathrm{FMR}_{\mathrm{sys}} = 1 - \prod_{i=1}^{m} \left[1 - \mathrm{FMR}_i * \sum_{j=Q-1}^{T-i} \binom{T-i}{j} \mathrm{FMR}_U^{\,j} (1 - \mathrm{FMR}_U)^{T-i-j}\right]^{N*P_i}. \tag{35}$$

Equation (35) holds regardless of the type of binning chosen. Note that the system false-match rate decreases with the decreasing system penetration coefficient. If ensemble binning is used, the penetration coefficients P_i are replaced with the single-penetration coefficient P_{ensemble}. Unlike the simple design used for development of Eq. (31), this ensemble-based design allows for reasonable system false-match rates even for systems with large N.

System Throughput

The final set of system equations is an approximation for the system-throughput rate S, which depends upon: 1) the hardware "one-to-one" comparison rate C; 2) the number of input samples compared to the database m; 3) the number of samples in the database N; and 4) the penetration coefficient, either computed on each sample P_i or over the ensemble P_{ensemble}.

We are assuming that the system-throughput rate is entirely limited by computational speed, not data collection time. This will be a fair assumption only for systems with large N. For systems with small N, throughput times will be limited by data collection speed, and other human factors.

Under the assumption that no matches will be found, the computational throughput rate S in customers per unit time can be written as

$$S = \frac{C}{\sum_{i=1}^{m} P_i * N} \tag{36}$$

where C is the hardware "one-to-one" computational rate. In the case where ensemble binning is used, Eq. (36) becomes.

$$S = \frac{C}{m * P_{\mathrm{ensemble}} * N}. \tag{37}$$

Violation of our assumption regarding binning independence increases the penetration coefficient and decreases throughput. Any matches found (false or correct) require additional comparisons over the remaining portion of the ensemble, further decreasing throughput, so Eqs. (36) and (37) are an optimistic upper bound.

This throughput rate must match the customer input on a time scale driven by operational requirements. Because of the various time units used, care must be taken in dimensional balancing when applying Eqs. (36) or (37).

It is generally true that hardware system costs increase with processing speed, C. Minimizing costs against a fixed customer throughput requirement pushes the system designer to decrease P_{sys}, through additional binning or filtering, thereby increasing false-nonmatch rate by Eq. (7), (25) or (29), and decreasing false-match rate by Eq. (35).

Examples

In this section we will apply the above equations to several types of biometric systems, specifically "one-to-one" "verification" systems, with and without a "three-strikes you're out" policy, and "one-to-many" and "M-to-many" "identification" systems.

"One-to-One" Systems

Consider a system in which a single sample is given and compared to a single enrolled template, perhaps contained on an identification card or associated with an enrolled user in a centralized database. The number of stored templates for each user is $T = 1$. If the templates are stored on a card, $N = T = 1$ and $P_{sys} = 1$. If the templates are in a centralized database, then $N = U * T$ and $P_{sys} = 1 / U$. In either case, $N * P_{sys} = T = 1$. There is no binning, so the bin-error rate is zero and the penetration coefficient is unity.

Using Eq. (21) with $m = 1$, we get $\mathrm{FNM}_{sys} = \mathrm{FNMR}$, which is as expected. Equation (25) could also be used, with $N = M = T = Q = 1$. We get

$$\mathrm{FNM}_{sys} = 1 - (1 - \mathrm{FNMR}) \cdot \sum_{j=0}^{0} \binom{0}{j} (1 - \mathrm{FNMR})^{0} (\mathrm{FNMR})^{0} (\mathrm{FNMR})^{0} = \mathrm{FNMR}.$$

Application of Eq. (29) also produces the same result.

The system false-nonmatch rate is most easily calculated with Eq. (31). We get $\mathrm{FNM}_{sys} = \mathrm{FNMR}$, which is expected. We could also calculate the system false-match rate using Eq. (35). We have

$$\mathrm{FMR}_{sys} = 1 - \prod_{i=1}^{1} \left[1 - \mathrm{FMR}_i * \binom{0}{0} \mathrm{FMR}_U^{0} (1 - \mathrm{FMR}_U)^{0} \right]^{1} = \mathrm{FMR}_i$$

as expected.

Applying Eq. (37) for throughput rate,

$$S = \frac{C}{m * P_{\mathrm{ensemble}} * N} = C$$

and we see that the throughput rate is exactly the single-sample-to-single-template hardware-comparison rate. This can be assumed to be so much faster than the time required for sample input that the throughput rate will be limited by human factors, not by hardware considerations.

"One-to-One" Systems with a "Three-Strikes" Decision Policy

Consider a system in which a single sample is given and compared to a single enrolled template, but the customer is given three tries to be identified. Any single match over the three tries results in a system "match" decision. The single-comparison error rates are assumed invariant over the match attempts. There is no binning, so the bin-error rate is zero and $N * P_{sys} = T = 1$, as in the previous example. By Eq. (20), FNM = FNMR. Using Eq. (21) with $m = 3$, we get

$$\mathrm{FNM}_{sys} = \prod_{i=1}^{m} \mathrm{FNMR}_i = \mathrm{FNMR}^3.$$

Equation (25) could also be applied, taking $Q = T = 1$ and $M = m = 3$. Because $Q = T = 1$, the requirement for $Q - 1$ matches against remaining templates has a probability of 1. Under these conditions, Eq. (25) yields

$$\mathrm{FNM}_{sys} = \prod_{i=1}^{m} [1 - (1 - \mathrm{FNM}_i)(1)] = \mathrm{FNMR}^3.$$

As previously noted, our computation above will underestimate the true false-nonmatch rate.

The false-match rate is computed using Eq. (31). Again, the probability of $Q - 1$ matches against remaining templates is 1.

$$\mathrm{FMR}_{sys} = 1 - \prod_{i=1}^{3} [1 - \mathrm{FMR}_i * 1]^{1} = 1 - [1 - \mathrm{FMR}]^3 = 3 * \mathrm{FMR} - O(\mathrm{FMR}^2)$$

where $O(\mathrm{FMR}^2)$ indicates terms on the order of the square of the false-match rate. The violation of the assumption of independence will cause the product in the above computation to be too large. Accordingly, this calculation will overestimate the true false-match rate.

These results for system false-match and false-nonmatch rate are identical to previously published results for the "three strikes" case [1].

By Eq. (37),

$$S = \frac{C}{m * P_{\text{ensemble}} * N} = \frac{C}{3}.$$

The throughput rate is one-third the hardware-comparison rate. Again, this is insignificant compared to the data-collection time. A trick that is usually employed to decrease the collection time is to collect and test the samples one at a time, collecting further samples only if a match is not determined.

"One-to-Several" Verification Systems

Now we will consider a system using only one biometric measure, but allowing several input samples and stored templates of varying presentations of that measure for each enrolled customer. If any input sample matches any of the enrolled templates, a system "match" results. As in the previous examples, $N * P_{\text{sys}} = T$, but here, $T > 1$. Similar to the development of Eq. (21), a false nonmatch occurs only if all comparisons of the m input samples to the T stored templates result in a false match. If the sample-to-template comparisons were independent, we could write

$$\text{FNM}_{\text{sys}} = \prod_{i=1}^{m} \text{FNM}_{i=1}^{T} = \text{FNM}^{m*T}.$$

Similarly, Eq. (31) could be rewritten as

$$\begin{aligned}\text{FMR}_{\text{sys}} &= 1 - \prod_{i=1}^{m} [1 - \text{FMR}_i]^{N*P_i} = 1 - \prod_{i=1}^{m} [1 - \text{FMR}_i]^{T} \\ &= m*T*\text{FMR} - O(\text{FMR}^2)\end{aligned}$$

where $O(\text{FMR}^2)$ indicates terms on the order of the square of the single-comparison false-match rate and the approximation is valid to the extent that this rate is small.

These equations indicate that the system false-nonmatch rate would go to 0 and the system false match rate to 1, as the number of total comparisons, $m * T$, increases. As previously noted, we have observed that a single-comparison false nonmatch increases the probability of subsequent nonmatches by the same customer within the same session. Additionally, we have operationally observed that the absence of a false match by a customer against a template decreases the probability of subsequent false match by that customer against the same template. Consequently, the development in this section overestimates the probability of a system false match and underestimates the probability for a system false nonmatch.

"One-to-Many" Single-Comparison Systems

Now we will consider a system in which a single sample is given and compared to a partitioned database of N individuals, enrolled with one template each. The system "match/nonmatch" decision is made on the basis of the single sample. In this case, N = large, $T = M = m = Q = 1$. This time there is individual sample binning, so the bin-error rate is nonzero and the penetration coefficient is less than 1.

Using data from the recent international automatic fingerprint identification system (AFIS) benchmark test [2], we will take values of performance-equation parameters that are consistent with large-scale fingerprint systems. Let's allow the penetration coefficient from endogenous binning to be $P = 0.5$ and apply gender-based filtering. Further, we will take the values $\varepsilon_{\text{BIN}} = 0.01$, $\text{FMR} = 10^{-5}$, and $\text{FNMR} = 0.05$ as obtainable by a general large-scale system.

We will first calculate the gender-based filter factor by applying Eq. (7). We will assume the population to be evenly divided between male and female and guess that no more than 2% of the population will be of unknown gender and that these unknowns will be equally male and female. The gender filter factor becomes by Eq. (7),

$$\begin{aligned}P_{\text{gender}} &= p_K + \sum_{i=1}^{K-1} (p_i + p_K) p_i \\ &= 0.02 + 0.51 * 0.49 + 0.51 * 0.49 \approx 0.51.\end{aligned}$$

Using Eq. (10), the total system penetration coefficient becomes

$$P_1 = \prod_{j=1}^{2} P_{1,j} = 0.5 * 0.51 = 0.26.$$

By Eq. (20),

$$\text{FNM}_1 = \text{FNMR}_1 + \varepsilon - \text{FNMR}_1 * \varepsilon \approx 0.06.$$

Using Eq. (25) to calculate the false-nonmatch rate for the system as

$$\begin{aligned}\text{FNM}_{\text{sys}} = \prod_{i=1}^{1} [1 - (1 - \text{FNM}_i) \\ \cdot \binom{0}{0} (1 - \text{FNM}_{i\cdot})^0 (\text{FNM}_{i\cdot})^0] = \text{FNM}_1 = 0.06\end{aligned}$$

we can see that the false-nonmatch rate is not directly dependent on N. However, as N increases, the system designer will be under increasing pressure to keep down the required computational rate by trading decreases in P_{sys} for increases in ε_{BIN}, and thereby increasing the false-nonmatch rate.

Now we consider the false-match rate as given by Eq. (31), which gives the probability that a sample will have one *or more* false matches over the $P_{\text{sys}} * N$ comparisons made. The expected number of system false matches $E[\text{FM}_{\text{sys}}]$ for a single sample over $P_{\text{sys}} * N$ comparisons is

$$E[\text{FM}_{\text{sys}}] = P_{\text{sys}} * N * \text{FMR}_{\text{sys}}. \tag{38}$$

Equation (38) comes directly from the expected value of a binomial distribution with parameters $n = P_{\text{sys}} * N$ and

$p = \Pr[\mathrm{FM}_{sys}]$. Equation (38) is interesting in its predictive value in estimating the formation of "candidate lists," the return of several false matches with each correct match as $E[\mathrm{FM}_{sys}]$ approaches and exceeds 1. Equation (38) tells us that candidate lists can be avoided only if $P_{sys} * N << 1 / \Pr[\mathrm{FM}_{sys}]$. This sets a natural limit for database size for the general biometric system as a function of the system penetration, single-comparison false-match rate, and system decision policy, if human intervention in the adjudication of candidate lists is to be avoided.

In our current example, with $P_{sys} = 0.26$ and $\mathrm{FMR} = 10^{-5}$, this implies that $N << 400{,}000$, if false matches are to be avoided. Let's take $N = 20{,}000$ as the size of our system.

Computing the false-match rate using Eq. (35) with $N = 20{,}000$,

$$\begin{aligned}\mathrm{FMR}_{sys} &= 1 - \prod_{i=1}^{1}\left[1 - \mathrm{FMR}_i * \binom{0}{0}\mathrm{FMR}_i^{0}(1-\mathrm{FMR}_i)^{0}\right]^{N*P_i} \\ &= 1 - [1 - \mathrm{FMR}]^{N*P_{sys}} \\ &\approx P_{sys} * N * 10^{-5} = 0.05\end{aligned}$$

with the approximation arising from the binomial expansion of the expression and being valid to the extent that $P_{sys} * N * \mathrm{FMR} << 1$. This indicates that, in such a system, approximately 5% of the customers would be falsely matched to one or more templates in the enrolled database.

Now, let's assume a customer input rate of N per year, or about 160 customers per working day, based on 250 working days per year. By Eq. (37), if the system throughput over an 8-hour period is to equal the customer input rate per day, then

$$S = \frac{C}{P_{\mathrm{ensemble}} * N} = \frac{C}{0.26 * 2 \times 10^{-4}} = \frac{160}{\mathrm{day}}.$$

The required hardware-comparison rate can be calculated as

$$C = \frac{160 * 0.26 * 2 \times 10^{-4}}{8\ \mathrm{hour\ day}} \approx \frac{9 \times 10^{5}}{3 \times 10^{4}\ \mathrm{sec}}$$
$$= 30\ \mathrm{comparisons\ /\ sec.}$$

"M-to-N" System Example

Now we will consider a specific instantiation of the M-to-N system in which four independent measures are used. (The Republic of the Philippines Social Security System identification project is using an ensemble of four fingerprints (both thumbs and both forefingers), with an initial search on the forefingers only. Confirmation of any matches is done against the remainder of the ensemble.) Both input samples and stored templates consist of a single ensemble of four measures. So, for this system $M = T = 4$ and $N = 4U$. We will consider a large-scale system where $U = 10^8$ so that $N = 4 \times 10^8$. An initial search will be made sequentially over two samples, so $m = 2$ and a total of $m * P_{sys} * N$ comparisons will be made. A "match" decision is made only if at least three input samples match the ensemble of a single enrolled individual, so $Q = 3$. We will use the same values for single-comparison error rates and penetration coefficient as in the above example. Gender-based filtering will again be used. It will be assumed that uniform bin-error rates, single sample penetration coefficients and single-comparison error rates apply to all measures in the ensemble.

Because we are using an ensemble of multiple, independent measures, we will have the choice of either binning on individual samples or the entire ensemble. It is interesting to compare the performance differences in the two approaches.

Regardless of approach used, we will partition the database by the four independent measures, placing all measures of each type into different, noncommunicating bins. For example, in a multiple fingerprint system, right thumb prints will be placed in one partition, left thumb prints in another. This is a filtering operation performed by the system operator at the time of data collection. Consequently, the inevitable errors in this procedure will not be considered in this analysis. Equation (3) applies and the penetration coefficient P_f owing to this filtering method is

$$P_f = \frac{1}{K} = \frac{1}{4} = 0.25.$$

Assume that we will use gender-based filtering and bin individually on each measure, using the penetration coefficients of the previous example. Then, by Eq. (10), the individual sample penetration coefficient would be

$$P_i = \sum_{j=1}^{3} P_{i,j} = 0.25 * 0.5 = 0.06.$$

If ensemble binning were used and if all the measures can be considered independent, by Eq. (11), the ensemble penetration coefficient would be

$$P_{\mathrm{ensemble}} = 0.25 * 0.51 * 0.5^{4} \approx 0.008.$$

If the partitionings of any of the measures are correlated, this value will be higher. We can see that ensemble binning decreases the penetration coefficient by nearly an order of magnitude over binning only on the individual samples. This translates into nearly an order of magnitude decrease in the hardware-computation rate for a fixed-throughput requirement.

A less obvious difference between systems using ensemble binning and systems using individual sample binning is in the system error rates.

Binning on individual samples, we can use Eq. (20) to write

$$\mathrm{FNM} = \mathrm{FNMR} + \varepsilon - \mathrm{FNMR} * \varepsilon \approx 0.06\ .$$

Using Eq. (25),

$$\mathrm{FNM}_{\mathrm{sys}} = \prod_{i=1}^{2}\left[1-(1-\mathrm{FNM})\sum_{j=2}^{4-i}\binom{4-i}{j}(1-\mathrm{FNM})^{j}(\mathrm{FNM})^{4-i-j}\right]$$
$$= [1-(1-\mathrm{FNM})(3(1-\mathrm{FNM})^{2}\mathrm{FNM} + 1(1-\mathrm{FNM})^{3})][1-(1-\mathrm{FNM})^{3}] \approx 0.01.$$

With ensemble binning, we calculate the ensemble binning error from Eq. (16) as

$$\varepsilon_{\mathrm{ensemble}} = 1-\prod_{i=1}^{4}(1-\varepsilon) \approx 0.04.$$

We use Eq. (27) to write

$$\mathrm{FNM} = \mathrm{FNMR} = 0.05.$$

Using Eq. (28) to calculate the system false-nonmatch rate, we find

$$\mathrm{FNM}_{\mathrm{sys}} = \varepsilon_{\mathrm{ensemble}} + [1-\varepsilon_{\mathrm{ensemble}}] \cdot \prod_{i=1}^{2}\left[1-(1-\mathrm{FNM})\sum_{j=2}^{4-i}\binom{4-i}{j}(1-\mathrm{FNM})^{j}(\mathrm{FNM})^{4-i-j}\right]$$
$$= \varepsilon_{\mathrm{ensemble}} - (1-\varepsilon_{\mathrm{ensemble}})[1-(1-\mathrm{FNM}) \cdot (3(1-\mathrm{FNM})^{2}\mathrm{FNM} + 1(1-\mathrm{FNM})^{3})] \cdot [1-(1-\mathrm{FNM})^{3}] \approx 0.05.$$

Thus, in this example, the system false-nonmatch rate using ensemble binning is five times that of using binning on individual samples.

Considering the false-match rate, Eq. (35) applies for both binning methods. The difference in its application is in the penetration coefficient used. For either method, Eq. (35) becomes

$$\mathrm{FMR}_{\mathrm{sys}} = 1 - \prod_{i=1}^{3}\left[1-\mathrm{FMR}*\sum_{j=2}^{4-i}\binom{4-i}{j}\mathrm{FMR}^{j}(1-\mathrm{FMR})^{2-i-j}\right]^{N*P_i}.$$

In this example we assume the penetration coefficient to be the same for all samples, even if individual sample binning is used. Therefore, the above equation can be rewritten as

$$\mathrm{FMR}_{\mathrm{sys}} = 1 - \left[\prod_{i=1}^{2}1-\mathrm{FMR}*\sum_{j=2}^{4-i}\binom{4-i}{j}\mathrm{FMR}^{j}(1-\mathrm{FMR})^{2-i-j}\right]^{N*P_i} \approx 1-\left[1-4*\mathrm{FNM}^{3}\right]^{N*P}.$$

If individual sample binning is used, $P = 0.06$ and the system false-match rate is approximately 1×10^{-7}, or one false match in every 10^{7} customer transactions. If ensemble binning is used, $P = 0.008$ and the system false-match rate is approximately 1×10^{-8}. With the ensemble binning method, the system has a smaller penetration coefficient, allowing fewer comparisons and fewer opportunities for a false match. We emphasize again that the above development assumed, without proof, statistical independence of all errors.

After the system is fully operational, with 10^{8} users enrolled, we will assume that renewals and re-issuances occur at a rate of about $1/5\ U$ per year. In a five-year period, we would expect about 10 false-match errors to occur with a system using individual sample binning and about 1 to occur with ensemble binning.

Based on the above, the input rate will be about 4×10^{5} customers per week, based on 50 working weeks per year. By Eqs. (36) and (37), the required hardware comparison rate for both individual sample and ensemble binning can be calculated as

$$S = \frac{C}{m*P*N} = \frac{4 \times 10^{5}}{\text{week}}.$$

Assuming a system availability of 20 hours per day, 7 days a week for the same 50 weeks per year, the required hardware-computational rate becomes

$$C = \frac{4 \times 10^{5} * 2 * 4 \times 10^{8} * P}{5 \times 10^{5}\ \text{sec}} \approx 6 \times 10^{8} * P\ \text{computations / sec.}$$

For individual sample binning with $P = 0.06$, $C = 4 \times 10^{6}$ computations per second. With ensemble binning, $P = 0.008$, and $C = 5 \times 10^{5}$ computations per second. Large-scale AFIS vendors are currently designing hardware systems with target processing rates on the order of a few hundred thousand comparisons per second.

Conclusions

In this article we derived equations for false-match and false-nonmatch error-rate prediction for the general M-to-N biometric identification system, under the simplifying, but limiting, assumption of statistical independence of all errors. For systems with large N, error rates were shown to be linked

to the hardware processing speed through the system penetration coefficient and the throughput equation. These equations are somewhat limited in their ability to handle sample-dependent decision policies and were shown to be consistent with previously published cases for "verification" and "identification" [1, 2]. Applying parameters consistent with the Philippine Social Security System benchmark test results for AFIS vendors [2], we established that biometric identification systems can be used in populations of 100 million people. Development of more generalized equations, accounting for error correlation and general sample-dependent thresholds, establishing confidence bounds, and substituting the inter-template for the impostor distribution under the template generating policy remain for future study.

Acknowledgments

The author is greatly indebted to Drs. John M. Colombi, Joseph P. Campbell, and Larry O'Gorman for their hours of careful reading of the text and their many helpful suggestions.

Keywords

Biometric identification, recognition error rates

References

[1] J.L. Wayman, "A scientific approach to evaluating biometric systems using a mathematical methodology," *Proc. CTST'97*, pp. 477-492.

[2] J.L. Wayman, "Benchmarking large-scale biometric system: issues and feasibility," *Proc. CTST Government'97*, Sept. 1997.

[3] D.C. Bright, "Examining the reliability of a hand geometry identity verification device for use in access control," Master's Thesis, Naval Postgraduate School, Monterey, CA, March 1987.

[4] M. Fuller, "Technological enhancements for personal computers," Master's Thesis, Naval Postgraduate School, Monterey, CA, March 1992.

[5] S.C. Geshan, "Signature verification for access control", Master's Thesis, Naval Postgraduate School, Monterey, CA, September 1991.

[6] D. Helle, "Examination of Retinal Pattern Threshold Levels and Their Possible Effect on Computer Access Control Mechanisms," Master's Thesis, Naval Postgraduate School, Monterey, CA, September 1985.

[7] G. Poock, "Fingerprint verification for access control," Naval Postgraduate School Report NPSOR-91-12, Monterey, CA, April 1991.

[8] G. Poock, "Voice verification for access control," Naval Postgraduate School Report NPSOR-91-01, Monterey, CA, October 1990.

[9] H. Kuan, "Evaluation of a biometric keystroke typing dynamics computer security system," Master's Thesis, Naval Postgraduate School, Monterey, CA, March 1992.

[10] L. Tirado, "Evaluation of fingerprint biometric equipment," Master's Thesis, Naval Postgraduate School, Monterey, CA, March 1991.

[11] Orkand Corporation, "Personal Identifier Report," California Department of Motor Vehicles, DMV 88-89, May 1990.

[12] J.P. Holmes, et al., "A performance evaluation of biometric identification devices," Sandia National Laboratories, SAND91-0276, June 1991.

[13] F. Bouchier, et al., "Laboratory evaluation of the IriScan prototype biometric identifier," Sandia National Laboratories, SAND96-1033, April 1996.

[14] P.J. Phillips, et al., "FERET (Face-Recognition Technology) recognition algorithm development and test results," Army Research Laboratory, ARL-TR-995, October 1996.

[15] P.J. Rauss, et al., "FERET (Face-Recognition Technology) recognition algorithms," *Proc. ATRWG Science and Technology Conference*, July 1996.

[16] A.K. Jain, et al., "An identity-authentication system using fingerprints," *Proc. IEEE*, vol. 85, no. 9, pp. 1365-1388, Sept. 1997.

[17] W. Shen, et al., "Evaluation of automated biometrics-based identification and verification systems," *Proc. IEEE*, vol. 85, no. 9, pp. 1464-1478, Sept. 1997.

[18] J.P. Campbell, "Speaker recognition: A tutorial," *Proc. IEEE*, vol. 85, no. 9, pg 1437-1462, Sept. 1997.

[19] J.L. Wayman, "The science of biometric technologies: Testing, classifying, evaluating," *Proc. CTST'97*, pp. 385-394.

[20] J.L. Wayman, "A generalized biometric identification system model," *Proc. IEEE Asilomar Conference on Signals, Systems, and Computers*, Nov. 1997.

[21] A. Higgins, et al., "Speaker verification using randomized phrase prompting," *Digital Signal Processing*, vol. 1, no.2, pp. 89-106, 1991.

[22] K. James and B. James, "NSA SAG Problem 97-25," monograph, UC Berkeley Dept. of Stats., to be published.

[23] D.W. Scott, *Multivariate Density Estimation*. New York: Wiley, 1992.

[24] J.L. Wayman, "Technical testing and evaluation of biometric devices" in Jain, et al, eds., *Biometrics: Information Security in a Networked Society*, Norwell, MA: Kluwer, 1998.

[25] B. Efron and R.J.Tibshirani, *An Introduction to the Bootstrap* New York: Chapman and Hall, 1993.

[26] P. Bickel, "NSA SAG Problem 97-2-1," monograph, UC Berkeley Department of Statistics, to be published.

[27] J.L. Hennessy and D.A. Patterson, *Computer Architecture: A Quantative Approach*, 2nd ed. San Francisco, CA: Morgan Stanley, 1996.

Jim Wayman is the Director of the U.S. National Biometric Test Center, located in the College of Engineering at San Jose State University in San Jose, California. The Test Center was established in 1995 under funding from the U.S. Government to advise government agencies on the use of biometric identification devices. Dr. Wayman received the Ph.D. degree in Engineering from the University of California at Santa Barbara in 1980 and joined the faculty of the Department of Mathematics at the U.S. Naval Postgraduate School in 1981. His early work was in acoustics and speaker recognition. In 1986, he became a full-time researcher for the Department of Defense in the areas of technical security and biometrics, inventing and developing a biometric system based on the acoustic resonances of the human head. Dr. Wayman holds two patents in speech processing and is the author of dozens of articles in technical journals and conference proceedings on biometrics, speech compression, acoustics, and network control. In addition to directing the Test Center, he teaches graduate and undergraduate courses in biometric identification at San Jose State University. He is a senior member of the IEEE.

Address for Correspondence: James L. Wayman, Biometric ID Research Director, U.S. National Biometric Test Center, College of Engineering, San Jose State University, One Washington Square, San José, CA, 95192. Tel: 408-924-4037. Fax: 408-924-3818. E-mail: jlwayman@aol.com.