TECHNICAL REPORT

ISO/IEC TR 24722

Second edition
2015-12-15

Information technology — Biometrics — Multimodal and other multibiometric fusion

Technologies de l'information — Biométrie — Fusion multimodale et autre fusion multibiométrique

Reference number
ISO/IEC TR 24722:2015(E)

Contents

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation on the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the WTO principles in the Technical Barriers to Trade (TBT) see the following URL: Foreword - Supplementary information

The committee responsible for this document is ISO/IEC JTC 1, *Information technology*, Subcommittee SC 37, *Biometrics*.

This second edition cancels and replaces the first edition (ISO/IEC/TR 24722:2007), which has been technically revised with the following changes:

— the original Clause 2 (Terminology issues) and Clause 7 (Scope and options for standardisation) are removed in this edition;

— Clause 2 (Terms and definitions) is aligned with ISO/IEC 2382-37;

— the current Clause 3, Clause 4, and Clause 5 have been technically revised in terminology, the state of arts updates, and other aspects. Such modifications have also been reflected in the bibliography.

Introduction

Some applications of biometrics require a level of technical performance that is difficult to obtain with a single biometric measure. Such applications include prevention of multiple applications for national identity cards and security checks for air travel. In addition, provision is needed for people who are unable to give a reliable biometric sample for some biometric characteristic types.

Use of multiple biometric measurements from substantially independent biometric sensors, algorithms, or characteristic types typically gives improved technical performance and reduces risk. This includes an improved level of performance where not all biometric measurements are available such that decisions can be made from any number of biometric measurements within an overall policy on accept/reject thresholds.

Of the various forms of multibiometric systems, the potential for multimodal biometric systems, each using an independent measure, has been discussed in the technical literature since at least 1974.[22][45] Advanced methods for combining measures at the score level have been discussed in Reference [15] and Reference [16]. At the current level of understanding, combining results at the score level typically requires knowledge of both genuine and impostor distributions. All of these measures are highly application dependent and generally unknown in any real system.

Research on the methods not requiring previous knowledge of the score distributions is continuing and research on fusion at both the image and feature levels is still progressing.

Given the current state of research into those questions and the highly application-dependent and generally unavailable data required for proper fusion at the score level, work on multibiometric fusion can, in the meantime, be considered mature. By intention, this Technical Report is not issued as an International Standard, in order not to force industrial solutions to conform to the methodology described herein. However, this Technical Report revision provides a mature technical description for developments of multibiometric systems. It will also provide a reference on multibiometric fusion for developers of other biometric standards and implementers.

Information technology — Biometrics — Multimodal and other multibiometric fusion

1 Scope

This Technical Report contains descriptions of and analyses of current practices on multimodal and other multibiometric fusion, including (as appropriate) references to more detailed descriptions.

This Technical Report contains descriptions and explanations of high-level multibiometric concepts to aid in the explanation of multibiometric fusion approaches including multi-characteristic-type, multiinstance, multisensorial, multialgorithmic, decision-level and score-level logic.

2 Terms and definitions

The following two categories of terms are defined here:

— terms that are specific to multimodal and multibiometric systems;

— terms that are not specific to multimodal and multibiometric systems, but are required to define the terms in the first category and not defined in the latest revision of ISO/IEC 2382-37.

For definitions of other terms in the subject field of biometrics, refer to ISO/IEC 2382-37. For the purposes of this document, the terms and definitions given in ISO/IEC 2382-37 and the following apply.

2.1
biometric data source
information channel (e.g. sensors, characteristic types, algorithms, instances or presentations) that is the origin of data (e.g. captured biometric sample, extracted features, comparison score, rank or decision) treated in fusion algorithms

2.2
biometric process
automated process using one or more biometric characteristics of a single individual for the purpose of enrolment, verification, or identification

2.3
biometric fusion
combination of information from multiple sources, i.e., sensors, characteristic types, algorithms, instances or presentations

2.4
cascaded system
system where pass/fail thresholds of biometric samples are used to determine if additional biometric samples are required to reach an overall system decision

2.5
layered system
system where individual biometric scores are used to determine the pass/fail thresholds of other biometric data processing

2.6
multialgorithmic
using multiple algorithms for processing the same biometric sample

2.7
multibiometric
uses multiple biometrics that can be combined at image, feature, score and/or decision level

Note 1 to entry: Multibiometric has five distinct subcategories: *multi-characteristic-type* (2.10), *multiinstance* (2.11), *multisensorial* (2.13), *multialgorithmic* (2.6) and *multipresentation* (2.12).

2.8
multibiometric process
biometric process (2.2) involving the use of *biometric fusion* (2.3)

2.9
multibiometrics
automated recognition of individuals based on their biological or behavioral characteristics and involving the use of *biometric fusion* (2.3)

2.10
multi-characteristic-type
multi-type
using information from multiple types of biometric characteristics

EXAMPLE Biometric characteristics types include: face, voice, finger, iris, retina, hand geometry, signature/sign, keystroke, lip movement, gait, vein, DNA, ear, foot, scent, etc.

2.11
multiinstance
using multiple biometric instances within one biometric characteristic type

EXAMPLE Iris (left) + Iris (right), Fingerprint (left index) + Fingerprint (right index).

2.12
multipresentation
using either multiple presentation samples of one instance of a biometric characteristic or a single presentation that results in the capture of multiple samples

EXAMPLE Several frames from video camera capture of a face image (possibly but not necessarily consecutive).

Note 1 to entry: Multipresentation biometrics is considered a form of *multibiometrics* (2.9), if fusion techniques are employed. Many fusion and normalisation techniques are appropriate to the integration of information from multiple presentations of the same biometric instance.

2.13
multisensorial
using multiple sensors for capturing samples of one biometric instance

EXAMPLE For face: infrared spectrum, visible spectrum, 2-D image, and 3-D image; for fingerprint: optical, electrostatic, and acoustic sensors.

2.14
sequential presentation
capturing biometric samples in separate capture events to be used for *biometric fusion* (2.3)

2.15
simultaneous presentation
capturing biometrics samples in a single capture event to be used for *biometric fusion* (2.3)

3 Overview of multimodal and other multibiometric systems

3.1 General

In general, the use of the terms multimodal or multibiometric indicates the presence and use of more than one characteristic type, sensor, instance, and/or algorithm in some form of combined use for making a specific biometric identification or verification decision. The methods of combining multiple samples, comparison scores or comparison decisions can be very simple or mathematically complex. For the purpose of this Technical Report, any method of combination will be considered a form of "fusion". Combination techniques will be covered in Clause 4.

Multimodal biometrics were first proposed, implemented and tested in the 1970s. Combining measures was seen as a necessary future requirement for biometric systems. It was widely thought that combining multiple measures could increase either security by decreasing the false acceptance rate or user convenience by decreasing the false rejection rate. These systems did not seem to advance into practical applications.

The use of fusion and related methods has been a key tool in the successful implementation of large-scale automated fingerprint identification systems (AFISs), starting in the 1980s. Until recently, multiple characteristic types have not been used in AFIS; however, most methods of fusion discussed elsewhere in this Technical Report have been successfully implemented using fingerprints alone. Some of the ways that fusion has been implemented in AFISs include the following:

— image (also known as sample) fusion in creating a single "rolled" image from a series of plain impressions on a livescan device;

— template fusion in the use of multiple feature extraction algorithms on each fingerprint image;

— multiinstance fusion in the use of fingerprints from all ten fingers;

— multipresentation fusion in the use of rolled and slap (plain) fingerprints;

— algorithm fusion for the purpose of efficiency (cost, computational complexity, and throughput rate); generally, comparators are used as a series of filters in order of increasing computational complexity. These are generally implemented as a mix of decision and score-level fusion;

— algorithm fusion for the purpose of accuracy (decreasing false accept rate and/or false reject rate, lessening sensitivity to poor-quality data); comparators are used in parallel, with fusion of resulting scores.

The use of fusion has made AFIS possible because of fusion's potential in improving both accuracy and efficiency.

Most work to date on multibiometrics has focused only on improving false acceptance and false rejection error rates. Some research work considers the use of multibiometrics to flexibly improve usability, security or accuracy.[64] Further, multibiometrics also aims at decreasing the overall failure-to-enrol rate (FTE) especially in biometric systems where user cooperation is not expected (e.g. video surveillance systems). Multibiometrics is an effort to produce a biometric decision even if only a subset of the expected biometric characteristics were captured.[66]

To further the understanding of the distinction among the multibiometric categories, Table 1 illustrates the basic distinctions among categories of multibiometric implementation. The key aspect of the category that makes it multi-"something" is shown in boldface.

Table 1 — Multibiometric categories illustrated by the simplest case of using 2 of something

Category	Characteristic type	Algorithm	Instance	Sensor
Multi-characteristic-type	2 (always)	2 (always)	2 (always)	2 (usually)[b]
Multialgorithmic	1 (always)	2 (always)	1 (always)	1 (always)
Multiinstance	1 (always)	1 (always)	2 (always)	1 (usually)[c]
Multisensorial	1 (always)	1 (usually)[a]	1 (always, and same instance)	2 (always)
Multipresentation	1	1	1	1

[a] It is possible that two samples from separate sensors could be processed by separate “feature extraction” algorithms, and then through a common comparison algorithm, making this “1.5 algorithms”, or two completely different algorithms.

[b] Exception: a multi-characteristic-type system with a single sensor used to capture two different characteristic types. For example, a high-resolution image used to extract face and iris or face and skin texture.

[c] Exception may be the use of two individual sensors to each capture one instance, for example, possibly a two-finger fingerprint sensor.

Multi-characteristic-type biometric systems take input from single or multiple sensors that capture two or more different types of biometric characteristics. For example, a single system combining face and iris information for biometric recognition would be considered a “multi-characteristic-type” system regardless of whether face and iris images were captured by different imaging devices or the same device. It is not required that the various measures be mathematically combined in anyway. For example, a system with fingerprint and voice recognition would be considered “multi-characteristic-type” even if the “OR” rule was being applied, allowing users to be verified using either of the characteristic types.

Multialgorithmic biometric systems receive a single sample from a single sensor and process that sample with two or more algorithms. This technique could be applied to any characteristic type. Maximum benefit (theoretically) would be derived from algorithms that are based on distinctly different and independent principles such as either features they extract from the biometric sample (e.g. finger minutiae versus finger pattern) or approaches to comparison (e.g. different algorithms comparing minutiae).

Multiinstance biometric systems use one (or possibly multiple) sensor(s) to capture samples of two or more different instances of the same biometric characteristic. For example, systems capturing images from multiple fingers are considered to be multiinstance rather than multi-characteristic-type. However, systems capturing, for example, sequential frames of facial or iris images are considered to be multipresentation rather than multiinstance.

Multisensorial biometric systems sample the same instance of a biometric characteristic with two or more distinctly different sensors. Processing of the multiple samples can be done with one algorithm, or some combination of multiple algorithms. For example, a face recognition application could use both a visible light camera and an infrared camera coupled with a specific frequency (or several frequencies) of infrared illumination.

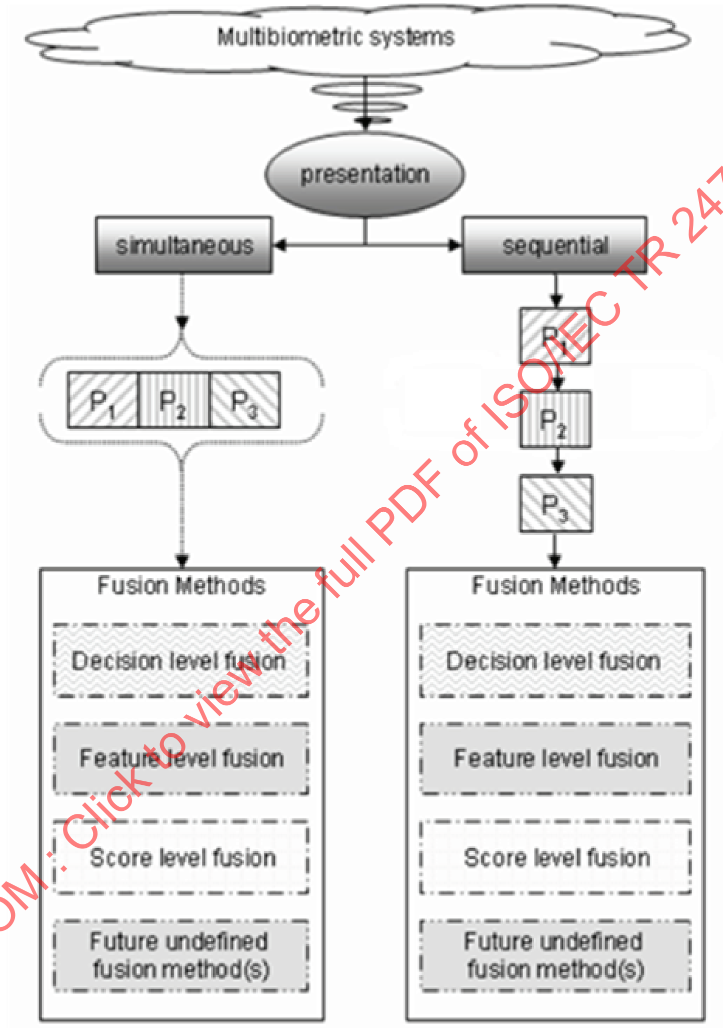
For a specific application in an operational environment, there are numerous system design considerations, and trade-offs that should be made, among factors such as improved performance (e.g. identification or verification accuracy, system speed and throughput, robustness, and resource requirements), acceptability, circumvention, ease of use, operational cost, environmental flexibility, and population flexibility.[40]

Especially for a large-scale human identification system, there are additional system design considerations such as operation and maintenance, reliability, system acquisition cost, life cycle cost, and planned system response to identified susceptible means of attack, all of which will affect the overall deployability of the system.[40]

3.2 Simultaneous and sequential presentation

3.2.1 General multibiometric system model

A general multibiometric system model is shown in Figure 1. For explanatory purposes, this model uses three biometric samples (P1, P2, P3) from three unique biometric characteristic types, except for where specified differently. At the topmost level, a subject presents their biometric characteristic(s) to the system. Dependent upon the system design, there are two methods of presenting characteristics for acquisition by the system: **simultaneous** and **sequential**.



NOTE The presentation (simultaneous or sequential) method induce or general different fusion process. The purpose of including this information is to illustrate considerations that can influence multibiometric system design.

Figure 1 — Multibiometric system model

3.2.2 Simultaneous presentation

Simultaneous presentation (with successful capture) provides biometric sample(s) from multiple characteristic types in a single event (e.g. a face and iris taken from the same camera). System designs that utilize simultaneous acquisition would tend toward high throughput applications at the expense of possible added complexity (to synchronize sample collection) or difficulty of use (dual sensor interaction, user multi-tasking).

3.2.3 Sequential presentation

Sequential capture acquires biometric sample(s) from one or multiple characteristic types in separate events. Sequential capture may be utilized in three concepts discussed in the literature. The first is multiinstance, which is the use of two or more instances within one characteristic type for a subject, i.e. Fingerprint (left index) + Fingerprint (right index). In this example, one single digit fingerprint reader is used twice in sequence. The second concept is multi-characteristic-type, which is the use of multiple different biometric characteristic types captured from one or more sensors for a subject, i.e. Hand + Face in sequence. The third concept is multisensorial, which is the use of two or more distinct sensors for capturing the same biometric feature(s) for a subject, but not at the same time. To avoid confusion with multi-characteristic-type, which may also capture biometric feature(s) from two or more distinct sensors, multisensorial can be clarified as “uni-characteristic-type multisensorial”. Examples for face recognition are infrared spectrum, visible spectrum, 2-D image, and 3-D image; for fingerprint recognition: optical, electrostatic and acoustic sensors.

3.3 Correlation

In multimodal biometric systems, the information being fused may be correlated at several different levels[56] as illustrated in the following examples.

— Correlation between characteristic types: This refers to biometric samples that are *physically related*, such as the speech and lip movement of a user.

— Correlation due to identical biometric samples: This is the case in multialgorithmic systems where the *same* biometric sample (e.g. a fingerprint image) or sub-sets of the biometric sample (e.g. voice, where an entire sample may be used by one algorithm and part of the sample by another) is subjected to different feature extraction and comparison algorithms (e.g. a minutiae-based comparator and a texture-based comparator).

— Correlation between feature values: A subset of feature values constituting the feature vectors of different characteristic types may be correlated. For example, the area of a user’s palm (hand geometry) may be correlated with the width of the face.

— Correlation among instances due to common operating procedures (e.g. common capture device and operator training).

— Correlation among instances due to subject behaviour (e.g. coloured contact lenses on both eyes).

However, in order to determine the *extent* of correlation, it is necessary to examine the comparison scores (or the ACCEPT/REJECT decision) pertaining to the comparators involved in the fusion scheme. In the multiple classifier system literature, it has been demonstrated that fusing uncorrelated classifiers leads to a significant improvement in comparison performance.[56]

For two classifiers of reasonable accuracy involved in a fusion scheme, score outputs from inputs that come from the same subject may, but need not, be correlated. Therefore it is more appropriate to consider the correlation of classifier errors as described by Reference [20]. The correlation ρ_{n_c} is given by Formula (1):

$$\rho_{n_c} = \frac{nN_c^f}{N - N_c^t - N_c^f + nN_c^f} \quad (1)$$

where

n is the number of classifiers under test;

N is the total number of sequences;

N_c^f is the number of sequences where all classifiers have an incorrect output at threshold C;

N_c^t is the number of sequences where all classifiers have a correct output at a threshold C.

NOTE This expression is relevant for computing the correlation of errors at the *decision* level.

4 Levels of combination

4.1 Overview

As a basis for the definition of levels of combination in multibiometric systems, we first introduce the single-biometric process and its building blocks, using the example of an authentication system. Figure 2 shows the block diagram of a single-biometric process.

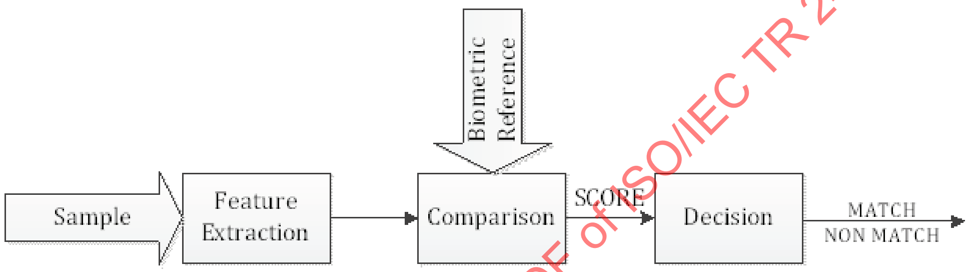


Figure 2 — Single biometric process (generic)

A biometric sample captured by a biometric sensor (e.g. a fingerprint image) is fed into the Feature Extraction module. Using signal processing methods, the feature extraction module converts a sample into Features (e.g. fingerprint minutiae), which form a representation apt for comparison. Usually, multiple features are collected into a feature vector. The Comparison module takes the feature vector as input and compares it to a Biometric Reference. The result is a comparison Score, which is used by the Decision module to decide (e.g. by applying a threshold) whether the presented sample matches with the stored template. The outcome of this decision is a binary *match* or *non-match*.

Generalizing the above process to multiple biometrics, there are several levels at which fusion can take place.

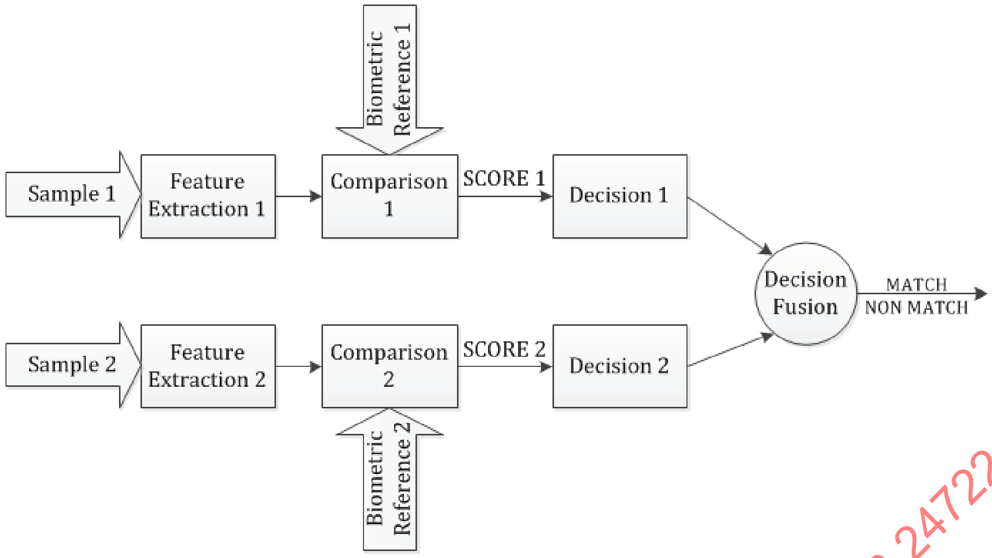
These include consolidating information at the (a) decision level, (b) comparison score level, (c) feature level, and (d) sample level. Note that fusion at levels (a) and (b) occur after the comparison module is invoked, while levels (c), and (d) occur before the comparator. Although integration is possible at these different levels, fusion at the feature set level, the comparison score level and the decision level are the most commonly used. Figure 3 illustrates the different levels of fusion for the case of a multimodal system.[7][41]

a) **Decision level:** each individual biometric process outputs its own Boolean result. The fusion process fuses them together by a combination algorithm such as AND and OR, possibly taking further parameters such as sample quality scores as input.

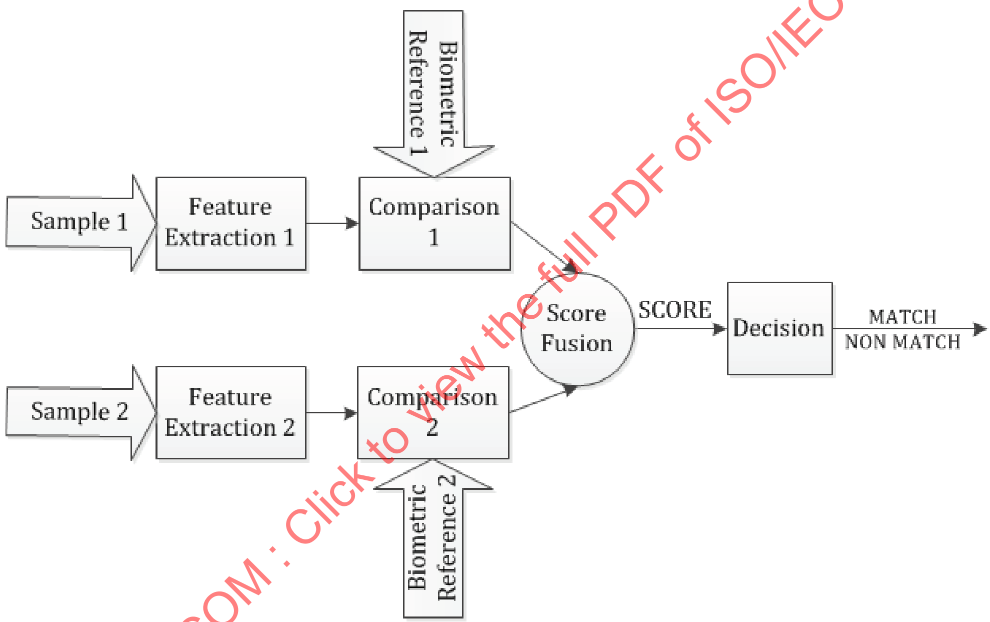
b) **Score level:** Each individual biometric process typically outputs a single comparison score but possibly multiple scores. The fusion process fuses these into a single score or decision, which is then compared to the system acceptance threshold.

c) **Feature level:** Each individual biometric process outputs a collection of features. The fusion process fuses these collections of features into a single feature set or vector.

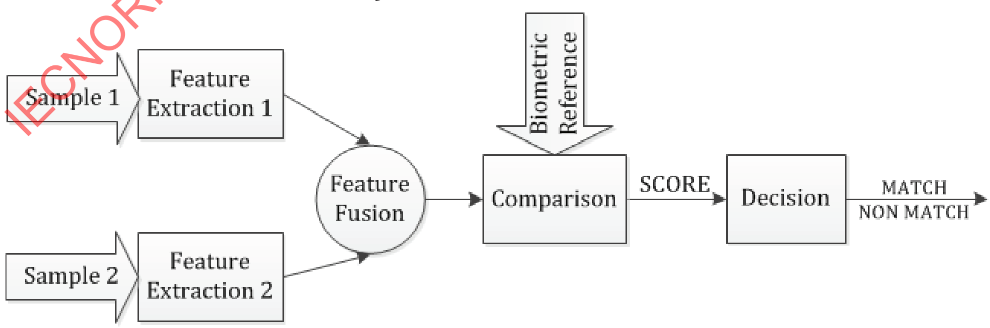
d) **Sample level:** Each individual biometric process outputs a collection of samples. The fusion process fuses these collections of samples into a single sample.



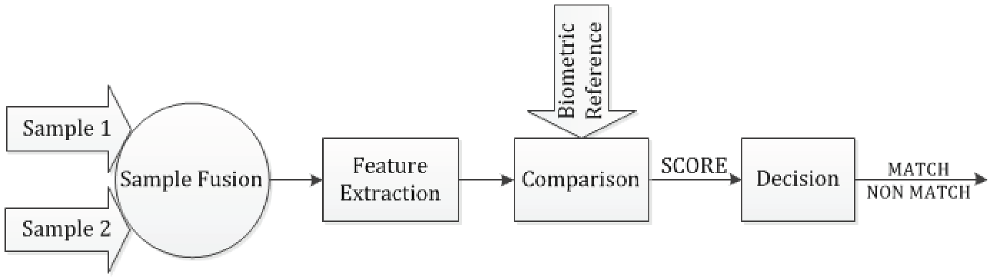
a) Decision-level fusion



b) Score-level fusion



c) Feature-level fusion



d) Sample-level fusion

NOTE Sample 1 and Sample 2 for c) may be the same sample.

Figure 3 — Different levels of fusion for the case of a multimodal system

For simultaneous or sequential biometric sample acquisition, features are extracted and are compared against the template. P1, P2, and P3 from Figure 1 refer to the comparison score from the comparison against the reference template. How the comparison scores are determined is system dependent and outside the scope of this Technical Report. The comparison scores of P1, P2, and P3 are then sent to the fusion module for a final result. In multibiometric systems, the fusion may occur at the decision or score level.

4.2 Decision-level fusion

4.2.1 Simple decision-level fusion

Decision-level fusion occurs after a comparison decision has been made for each biometric component. It is based on the binary result values *match* and *non-match* output by the decision modules (see Figure 3 a), Decision-level fusion).

For biometric systems composed of a small number of components, it is convenient to assign logical values to comparison outcomes so that fusion rules can be formulated as logical functions. The behaviour of the two most widely used functions, AND and OR, are listed in Table 2, assuming a pair of decision-level outputs.

Table 2 — AND and OR fusion of decisions for a case of two biometric characteristic types

Decision Biometrics 1	Decision Biometrics 2		AND-fused decision	OR-fused decision
X	X		X	X
X	•		X	•
•	X		X	•
•	•		•	•

X Non-match
• Match

For biometric systems using many components, voting schemes have been established as fusion rules, the most common of which is majority voting rule. The AND and OR are specific examples of voting schemes.

4.2.2 Advanced decision-level fusion

4.2.2.1 General model

Decision-level fusion is based upon individual accept/reject decisions for each sample. The two sub groups of advanced decision-level fusion are **layered** and **cascaded**. A layered system features with adjustable thresholds computed by using individual biometric scores to determine the pass/fail thresholds for other biometric data processing. A cascaded system features with fixed thresholds is pass/fail thresholds of characteristic type-specific biometric samples to determine if additional biometric samples from other characteristic types are required to reach an overall system decision. Decision-level fusion for the two subgroups is shown in Figure 4.

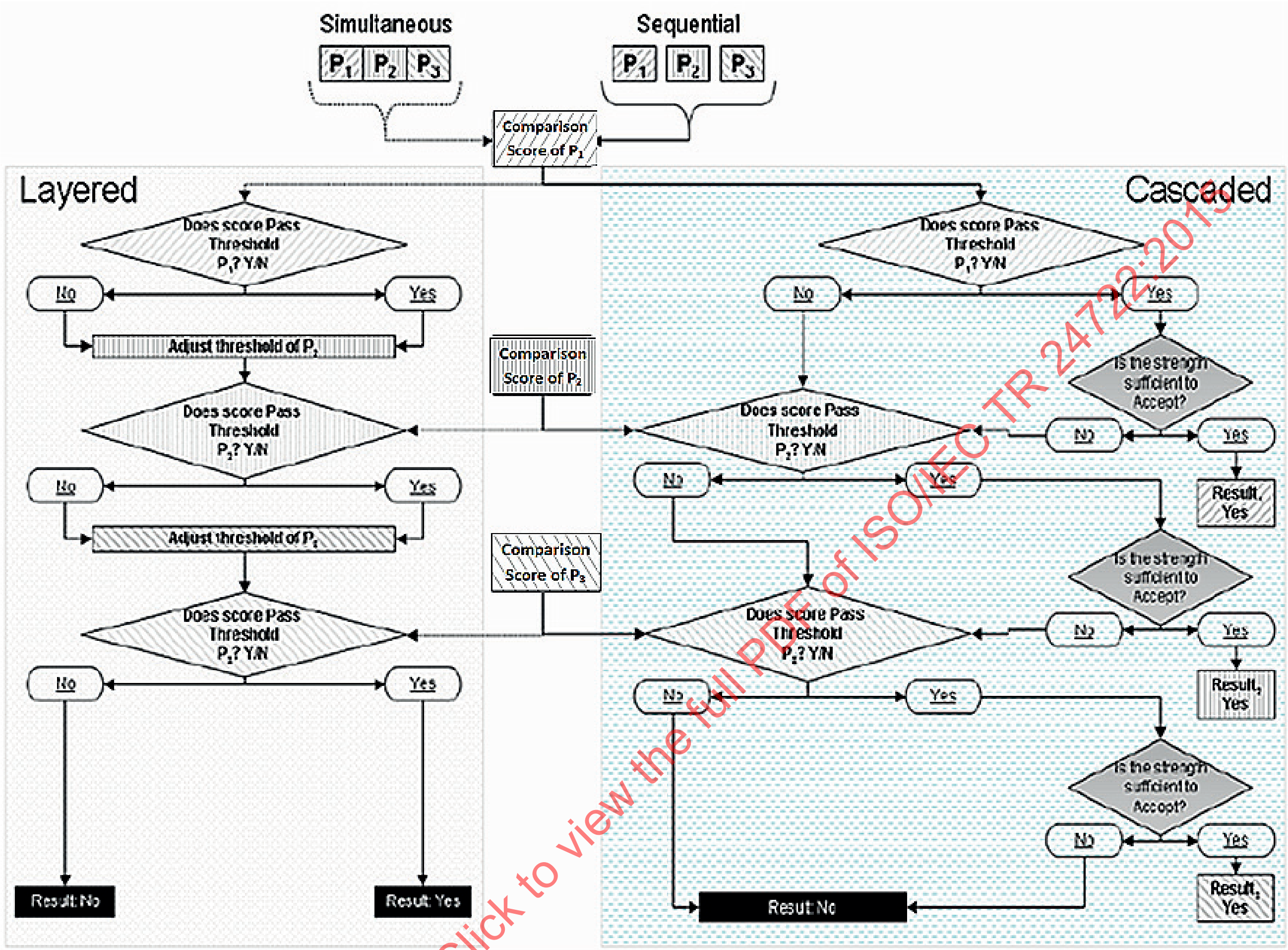


Figure 4 — Advanced decision-level fusion

4.2.2.2 Layered system

Independent of whether the presentation was simultaneous or sequential, the comparison score of P1 enters the layered system. The system processes the score against the system defined threshold. If it passes the criteria/threshold for characteristic type P1, the output would adjust (raise or lower) the threshold needed to pass for characteristic type P2. If P1 fails to meet the criteria/threshold for characteristic type P1, then the output most likely would increase the threshold required for characteristic type P2. Upon completion of processing P1 and resetting the thresholds requirements for characteristic type P2, the comparison score of P2 enters the system. The process iterates as discussed above for P2 and P3. Once the characteristic type P3 process is completed, a final accept/reject decision is made.

4.2.2.3 Cascaded system

Independent of simultaneous or sequential presentation, cascaded systems rely on at least one biometric sample.

If the first sample does not meet the requirements, additional samples are compared. Using Figure 4 as the model for this discussion, comparison score P1 enters the system and is compared against the threshold for sample P1. If the score exceeds the criteria/threshold required for P1, a subsequent decision is made on the strength of the result (which could also include sample quality measures). If this strength is sufficient, the subject is accepted. If the score of P1 fails the initial threshold test or passes the initial threshold test but fails the strength decision, cascaded systems require the use of the score of P2. This process is repeated for scores P2 and P3. Note that cascaded systems might not require P2 or P3 to be captured if P1 passes the threshold and strength test.

4.3 Score-level fusion

4.3.1 Overview

In score-level fusion, each system provides comparison scores indicating the proximity of the feature vector with the Biometric Reference vector. These scores can then be combined to improve the comparison performance.

From a theoretical point of view, biometric processes can be combined reliably to give a guaranteed improvement in comparison performance. Any number of suitably characterized biometric processes can have their comparison scores combined in such a way that the multibiometric combination is guaranteed (on average) to be no worse than the best of the individual biometric devices. The key is to identify correctly the method which will combine these comparison scores reliably and maximize the improvement in comparison performance.

The mechanism (for this sort of good combination of scores within a multibiometric system) shall follow at least two guidelines. Firstly, each biometric process shall produce a score, rather than a hard accept/reject decision, and make it available to the multibiometric combiner. Secondly, in advance of operational use, each biometric process shall make available to the multibiometric combiner, its technical performance (such as score distributions) in the appropriate form (and with sufficient accuracy of characterisation).

Both verification (1:1) and identification (1:N) systems can support fusion at the comparison score level. However, identification systems can also integrate information available at the rank level (which is a form of score level with multiple scores or indices based on scores). In identification systems, a template from a biometric sample is compared against templates from a subset of identities present in the database and, therefore, a sequence of ordered comparison scores pertaining to these identities is available. Reference [23] describes three methods to combine the ranks assigned by the different comparators. In the *highest rank method*, each possible match is assigned the highest (minimum) rank as computed by different comparators. Ties are broken randomly to arrive at a strict ranking order and the final decision is made based on the combined ranks. The *Borda count* method uses the sum of the ranks assigned by the individual comparators to calculate the combined ranks. The *logistic regression* method is a generalization of the Borda count method where the weighted sum of the individual ranks is calculated and the weights are determined by logistic regression.

4.3.2 Score normalization

Score normalisation methods attempt to map the scores of each biometric process to a common domain. Some approaches are based on the Neyman-Pearson lemma, with simplifying assumptions. For example, mapping scores to likelihood ratios allows them to be combined by multiplying under an independence assumption. Other approaches may be based on modifying other statistical measures of the comparison score distributions.

The parameters used for normalisation can be determined using a fixed training set or adaptively based on the current feature vector. The computed characteristic may represent only "estimates" of the underlying population characteristics. Score normalisation is closely related to score-level fusion since it affects how scores are combined and interpreted in terms of biometric performance. As in Reference [32]:

a) The comparison scores at the output of the individual comparators *may not be homogeneous*. For example, one comparator may output a distance (dissimilarity) measure while another may output a similarity measure.

b) Further, the outputs of the individual comparators *need not be on the same numerical scale* (range).

c) Finally, the comparison scores at the output of the comparators *may follow different statistical distributions*.

Due to these reasons, scores are generally normalized prior to fusion into a common domain. Figure 5 depicts a score-level fusion framework for processing two biometric samples, taking normalisation into account.

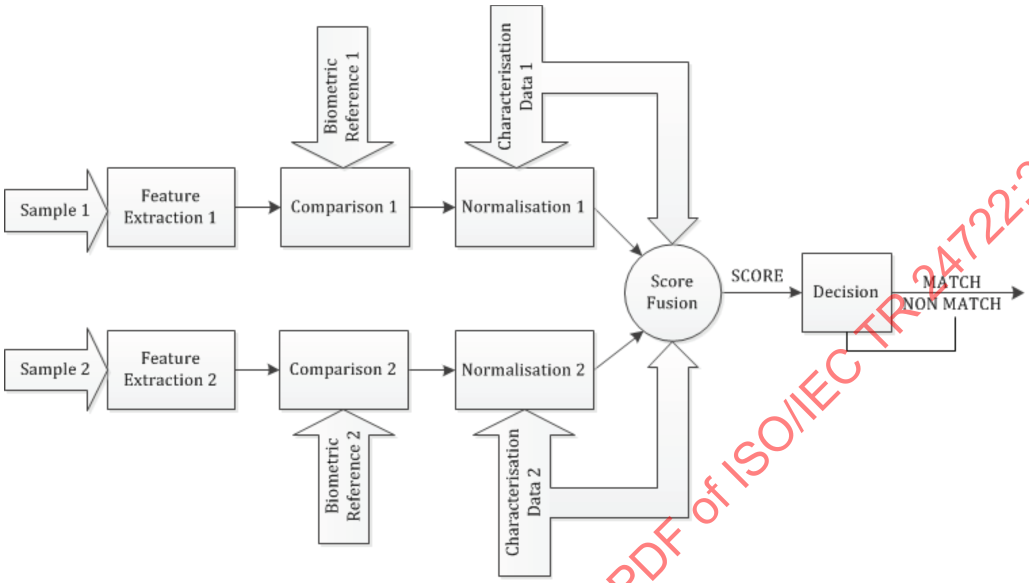


Figure 5 — Framework for score-level fusion

Table 4 lists, under the framework of Figure 5, several commonly used score normalisation methods. Note that some fusion methods use probability density functions (PDFs) directly and do not require normalisation methods.

Table 3 defines the symbols used in Table 4. In some cases, PDFs are used to convert raw/native scores directly into Probability of False Accept and thus to a decision without need to have native scores brought to a common reference range using normalization.

Table 3 — Symbols used for score normalisation formulas

Statistical measures	Characterisation data		
	Genuine distribution	**Impostor distribution**	**Both genuine and impostor distributions**
Minimum score	$S^G{}_{Min}$	$S^I{}_{Min}$	$S^B{}_{Min}$
Maximum score	$S^G{}_{Max}$	$S^I{}_{Max}$	$S^B{}_{Max}$
Mean score	$S^G{}_{Mean}$	$S^I{}_{Mean}$	$S^B{}_{Mean}$
Median score	$S^G{}_{Med}$	$S^I{}_{Med}$	$S^B{}_{Med}$
Score standard deviation	$S^G{}_{SD}$	$S^I{}_{SD}$	$S^B{}_{SD}$
Constant	**C**	**C**	**C**
Probability density fusion	PDF^G	PDF^I	N.A.
Centre of PDF crossover	S_{center}		
Width of PDF crossover	S_{width}		

S Similarity score.

G Genuine.

I Impostor.

B Both.

Table 4 — Examples of score normalisation methods

Method	Formula	Data elements	Comment
Min-max (MM)	$S' = \left(S - S^B{}_{Min}\right) \big/ \left(S^B{}_{Max} - S^B{}_{Min}\right)$	$S^B{}_{Min}$ $S^B{}_{Max}$	— Uses empirical data (or theoretical limit or vendor provided) — No accounting for nonlinearity
Z-score	$S' = \left(S - S^I{}_{Mean}\right) \big/ S^I{}_{SD}$	$S^I{}_{Mean}$ $S^I{}_{SD}$	— Assumes normal distribution — Symmetric about mean — Assumes stability of both distributions across populations
Median absolute deviation (MAD)	$S' = \left(S - S^B{}_{Med}\right) \big/ \left(C \cdot \mathrm{median}\left\lvert S - S^B{}_{Med}\right\rvert\right)$	$S^B{}_{Med}$ C	— Assumes stability of both distributions across populations
Hyperbolic tangent (Tanh)	$S' = 0{,}5\left(\tanh\left(C\left(S - S^G{}_{Mean}\right) \big/ S^G{}_{SD}\right) + 1\right)$	$S^G{}_{Mean}$ $S^G{}_{SD}$	— Mean and variance of transformed data distribution — Assumes stability of both distributions across populations

NOTE This table lists two types of normalisation schemes: (a) schemes that modify the location and scale parameters of the score distribution and (b) schemes that consider only the overlap region of the genuine and impostor scores. Thus, the min-max, z-score, MAD and tanh techniques fall under category (a), while QQ and QLQ fall under category (b). Typically, category (b) techniques are used *after* having applied one of the category (a) schemes.

a Refer to Reference [62].

Table 4 *(continued)*

Method	Formula	Data elements	Comment
Adaptive (AD)[a] a) Two-quadrics (QQ) b) Logistic c) Quadric-line-quadric (QLQ)	$$n_{AD} = \begin{cases} \frac{1}{c} n_{MM}^2, & n_{MM} \le c \\ c + \sqrt{(1-c)(n_{MM}-c)}, & \text{otherwise} \end{cases}$$ $$n_{AD} = \frac{1}{1 + A \cdot e^{-B \cdot n_{MM}}}$$ $$n_{AD} = \begin{cases} \frac{1}{c} n_{MM}^2, & n_{MM} \le \left(c - \frac{w}{2}\right) \\ n_{MM}, & \left(c - \frac{w}{2}\right) < n_{MM} \le \left(c + \frac{w}{2}\right) \\ \left(c + \frac{w}{2}\right) + \sqrt{\left(1 - c - \frac{w}{2}\right)\left(n_{MM} - c - \frac{w}{2}\right)}, & \text{otherwise} \end{cases}$$	c w Δ $A = \frac{1}{\Delta} - 1$ $B = \frac{\ln A}{c}$	— Assumes non-linearity — Three modelling methods — Assumes stability of both distributions across populations — n_{AD} = adaptive normalisation score; n_{MM} = min-max normalized score; c = center of overlap of genuine and impostor score distributions; w = width of the overlap; Δ = a small value (0,01 in Reference [62])
Biometric Gain against Impostors (BGI)	$P_{Si\|I} / P_{Si\|G}$, $\quad P_{Si\|G}$ = Value of PDF^G at score S_i $P_{Si\|I}$ = Value of PDF^I at score S_i	PDF^G PDF^I	— Assumes stability of both distributions across populations
BioAPI	$S' = FAR_{(\text{threshold}=\text{score})}$	PDF^I	— Assumes stability of imposter distribution
Borda count	N – Rank (S) where N is the number of alternatives.	Rank	— Applicable only to 1:N comparison

NOTE This table lists two types of normalisation schemes: (a) schemes that modify the location and scale parameters of the score distribution and (b) schemes that consider only the overlap region of the genuine and impostor scores. Thus, the min-max, z-score, MAD and tanh techniques fall under category (a), while QQ and QLQ fall under category (b). Typically, category (b) techniques are used *after* having applied one of the category (a) schemes.

[a] Refer to Reference [62].

4.3.3 Score fusion methods

When individual biometric comparators output a set of possible matches along with the quality of each match (comparison score), integration can be done at the comparison score level. This is also known as fusion at the measurement level or confidence level. The comparison score output by a comparator contains the richest information about the input biometric sample in the absence of feature-level or sensor-level information. Furthermore, it is relatively easy to access and combine the scores generated by several different comparators. Consequently, integration of information at the comparison score level is the most common approach in multimodal biometric systems. Table 5 provides an outline of several score fusion methods and their associated needs for data that characterise the comparator performance.

NOTE This is valid only in the case where a rank and/or a comparison score is/are available for all references present in the set of possible matches given by each algorithm.

In the context of verification, there are two distinct approaches to score-level fusion. One approach is to formulate it as a classification problem, while the other approach is to treat it as a combination problem.[32][35] In the classification approach, a feature vector is constructed using the comparison scores output by the individual comparators; this feature vector is then classified into one of two classes: "Accept" (genuine user) or "Reject" (impostor). Generally, the classifier used for this purpose (e.g. decision tree, neural network, support vector machine, K-nearest neighbour, random forest, etc.) is capable of learning the decision boundary irrespective of how the feature vector is generated.[6][64][65] Hence, the output scores of the different characteristic types can be non-homogeneous (distance or similarity metric, different numerical ranges, etc.) and no processing is required prior to presenting them to the classifier. In the combination approach, the individual comparison scores are combined to generate a single scalar score, which is then used to make the final decision.[38] To ensure a meaningful combination of the scores from the different characteristic types, if necessary, the scores may be first transformed to a common domain prior to combining them. This is known as score normalisation (as discussed in 4.3.2).[27]

As part of a pattern classification problem, in the classification approach, the fusion module design aims at finding an optimal two-class classifier for genuine and impostor classes. The classifier uses the vector of comparison scores provided by the comparators and assigns one of the two classes to it. For this purpose, the classifier defines two decision regions in the feature vector space: one for genuine class and one for impostor class. These regions are separated by decision boundaries, which need to be optimized during the design of the fusion module. These decision boundaries can have various forms depending upon the complexity and the nature of the distributions of the two classes. They can be as simple as a line as in linear discriminant functions or more complex as in multilayer neural networks and support vector machines. The boundaries can also be determined from statistics such as the Neyman-Pearson likelihood ratio. Regardless of the chosen technique, the ultimate goal is to find decision boundaries that improve classification performance to fit the application.

Combination approaches are some of the simplest and most effective methods for biometric fusion, provided scores are homogeneous or can be normalised to make them so. Because of this simplicity and effectiveness, they are some of the most common methods for use in multibiometric systems. Kittler's theoretical framework for combining classifiers[38] describes some of the most popular techniques, these being the product, sum, max, min and median rules. Each of these techniques uses simple arithmetic or rule operations to combine scores from multiple sources. These techniques were extended by Reference [1] to allow weighting of the comparison scores based on performance. If more information on the distribution of comparison scores is available, then one may use Bayesian statistics in combining the scores of different biometric comparators as demonstrated by Reference [3]. Their technique takes into account the estimated accuracy of the individual classifiers during the fusion process. In general, fusion can be accomplished using a Bayesian classifier when sufficient training data is available. Let $P_i(S|G)$ and $P_i(S|I)$ denote the probability densities of score S (corresponding to the i^{th} characteristic type) under the genuine and impostor hypothesis, respectively. A simple Bayesian classifier (SBC) would make a MATCH/NO-MATCH decision based on the posterior densities P(G|S1, S2, ...SN) and P(I|S1, S2, ...SN). In the absence of sufficient training data (i.e. genuine and impostor comparison scores), it is not possible to reliably estimate the *joint density* involving multiple characteristic types. Thus, the posterior probability could be estimated by the *product* of individual densities, i.e., $P(G|S1, S2, ...SN) = \Pi P_i(S_i|G)$ and $P(I|S1, S2, ...SN) = \Pi P_i(S_i|I)$.

Table 5 — Examples of score fusion methods

Method	Score fusion equation	Characterisation data required					
		None	PDF_G	PDF_I	EER	V_G,V_I	Personal
Simple sum	$\sum\left(i = 1 \text{ to N}\right) S_i'$	o					
Minimum score	$\min\left(i = 1 \text{ to N}\right) S_i'$	o					
Maximum score	$\max\left(i = 1 \text{ to N}\right) S_i'$	o					
Comparator weighting	$\sum\left(i = 1 \text{ to N}\right) W_i \cdot S_i'$				o		
Comparator weighting with PDF fusion for decision[a]	$\sum\left(i = 1 \text{ to N}\right) W_i' \cdot S_i'$		o	o			
User weighting	$\sum\left(i = 1 \text{ to N}\right) W_i^* \cdot S_i'$						o

Table 5 *(continued)*

Method	Score fusion equation	Characterisation data required					
		None	PDF_G	PDF_I	EER	V_G,V_I	Personal
Weighted product	$\Pi\left(i = 1 \text{ to } N\right) W_i \cdot S_i{}'$				0		
Sum of probabilities Genuine	$\sum\left(i = 1 \text{ to } N\right) P_{G\|Si}$		0				
Sum of probabilities Impostor	$\sum\left(i = 1 \text{ to } N\right) P_{I\|Si}$			0			
Product of probabilities Genuine	$\Pi\left(i = 1 \text{ to } N\right) P_{G\|Si}$		0				
Product of probabilities Impostor	$\Pi\left(i = 1 \text{ to } N\right) P_{I\|Si}$			0			
BGI[b]	$\Pi\left(i = 1 \text{ to } N\right) BGI_i$		0	0			
Likelihood ratio[c]	PDF_G / PDF_I		0	0			
K-nearest neighbour						0	
Decision trees						0	
Support vector machines						0	
Discriminant analysis						0	
Neural network						0	

The following symbols and abbreviations are used in the table:

i — i-th biometric score

N — number of fusion inputs

$S_i{}'$ — i-th normalized match score

W_I — i-th matcher weight factor

$W_i{}^*$ — i-th user weight factor

$W_i{}'$ — i-th natcher weight factor in case of PDF fusion

BGI — biometric gain against impostors

PDF_G — probability density functions of scores from genuine users for each dimension

PDF_I — probability density functions of scores from impostors for each dimension

EER — equal error rate

V_G — N-dimensional genuine score vector, N is the number of modalities

V_I — N-dimensional impostor score vector, N is the number of modalities

$P_{G|Si}$ — value of PDF_G at score S_I

$P_{I|Si}$ — value of PDF_I at score S_I

[a] Refer to Reference [64].

[b] Refer to References [60] and [61].

[c] Refer to Reference [51].

4.4 Feature-level fusion

In feature-level combination, biometric information is fused after feature extraction but before comparison [see Figure 3 c)]. There are several ways features can be combined. The simplest form is to integrate the feature vectors (or sets if there is no implicit correspondence) of component biometrics and to apply feature classification methods to the combined feature vector. Where features from contributing multibiometrics are not independent, good feature-level combination should, in some circumstances, allow dependencies to be more fully exploited than by solely using score-level combination. Feature normalization is normally used before combining the real valued features (especially in case of feature concatenation). However, in case of binary features fusion, feature normalization is not used. This should give better overall performance. However, fusion at this level is difficult to achieve in practice because the feature vectors of multiple characteristic types might be incompatible (e.g. minutiae set of fingerprints and Eigen-coefficients of face), the relationship between the feature spaces of different biometric systems might not be known, concatenating two feature vectors might result in a feature vector with very large dimensionality leading to the “curse of dimensionality” and a significantly more complex comparator might be required in order to operate on the concatenated feature vector.[55]

Notwithstanding these challenges, fusion at the feature level has been attempted in several contexts. Reference [5] demonstrates feature-level fusion of face and ear characteristic types showing significant improvements in performance. Reference [41] integrates the palm-print and hand geometry features of an individual in order to enhance comparison performance. In their experiments, fusion at the comparison score level was observed to be superior to fusion at the feature level. However, Reference [55] combines the hand and face characteristic types of a user (multibiometrics), as well as the R, G, B channels of the face image of a user (multisensorial) at the feature level and demonstrate that a feature selection scheme may be necessary to improve comparison performance at this level. Thus, it is imperative that an appropriate feature selection scheme is used when combining information at the feature level.

Features can also be combined in a more complex way on an algorithmic level through co-registration. Most feature extraction algorithms require the localization of landmarks in order to establish a common coordinate frame between samples for feature extraction. In multibiometric systems, individual components can exchange landmarks or mutually support their extraction. This technique, called co-registration, is considered a form of feature-level combination. For example, a face recognition algorithm may provide eye locations for an iris recognition algorithm or depth landmarks in a 3D face recognition system may be used to correct the pose of faces in texture images.

5 Characterisation data for multibiometric systems

5.1 Overview

One of the most important aspects of normalisation and combination for multibiometric systems is the origin of parameters for such normalisation and/or combination. In the case of score-level combination using statistical pattern recognition theory, the PDFs of genuine and impostor score distributions are required. In other score-level combination and in feature-level and decision-level combination, there are usually important parameters that, in many cases, are required to be derived from characterisation data. Thus, this issue is all pervading and conditions the relevance of theoretical analysis of the optimal fusion rule.

This Clause is allocated to analysis and discussion of characterisation data, its expected origin(s), extent of its validity (e.g. through small sample sizes or other limitations on characterisation sample populations) and how such data would be disseminated or otherwise made available for use.

5.2 Use of characterisation data in normalisation and fusion

Score-level fusion combines the similarity scores from one or more comparators. In the multi-characteristic-type and multialgorithmic case, there will generally be two or more such comparison systems. In the multisensor, multiinstance, and multipresentation cases, only one comparator will usually be in use, but in any case, multiple scores will be available to a fusion module. The distribution of comparison scores will depend on the comparison system and the statistics of these variables will not usually be on any common range. Thus, the normalisation process of 4.3.2 is a necessary precursor of the fusion process.

The characterisation data, discussed in this subclause, is needed to support normalisation and fusion. At its most simple, this might be just the location and shape parameters of each score's "natural" distribution. For example, a face and fingerprint fusion scheme would use some prior estimates of the median and median absolute deviation (see Table 5) to effect normalisation of two scores. More usefully, a full specification (approximated) of the distribution of the scores would be used and such a description would be provided for both the genuine and impostor distributions.

Thus, a biometric system's characterisation data is just some representative summary of the statistics of its output scores. One powerful and simple characterization is the cumulative distribution function (cdf), which may be expressed as N pairs of (S_i, cdf(S_i)) or some functional fit of the data (see References [18] and [36]).

In score normalization-based processes, fusion is preceded by a transformation of each score to a common domain. The fusion information format defined in ISO/IEC 29159-1 is intended to flexibly support any of the popular transformations. By establishing a standardized means of data exchange, ISO/IEC 29159-1 supports a modular approach to biometric systems integration in which both the comparison and fusion algorithms remain protected as pieces of intellectual property. Thus, ISO/IEC 29159-1 envisages an application in which two (or more) underlying acquisition and comparison technologies (hand geometry and fingerprint, for example) each generate a score which is fed to a fusion module which has been initialized with an appropriate instance of the Fusion Information Format defined in ISO/IEC 29159-1.