FINAL
DRAFT

INTERNATIONAL
STANDARD

**ISO/IEC
FDIS
19795-1**

ISO/IEC JTC **1**/SC **37**

Secretariat: ANSI

Voting begins on:
2021-02-16

Voting terminates on:
2021-04-13

Information technology — Biometric performance testing and reporting —

Part 1:
Principles and framework

Technologies de l'information — Essais et rapports de performance biométriques —
Partie 1: Principes et canevas

Reference number
ISO/IEC FDIS 19795-1:2021(E)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword ..... **vi**
Introduction ..... **vii**
1 Scope **1**
2 Normative references **1**
3 Terms and definitions **1**
4 Abbreviated terms **5**
5 Conformance **6**
6 General biometric system **6**
6.1 Conceptual representation of general biometric system 6
6.2 Conceptual components of a general biometric system 7
6.2.1 Data capture subsystem 7
6.2.2 Transmission subsystem 7
6.2.3 Signal processing subsystem 7
6.2.4 Data storage subsystem 7
6.2.5 Comparison subsystem 7
6.2.6 Decision subsystem 8
6.2.7 Administration subsystem 8
6.2.8 Interface to external application 8
6.3 Functions of general biometric system 9
6.3.1 Enrolment 9
6.3.2 Verification of a positive biometric claim 9
6.3.3 Identification 10
6.4 Enrolment, verification and identification transactions 11
6.5 Performance measures 12
6.5.1 Error rates 12
6.5.2 Throughput rates 12
6.5.3 Types of performance testing 13
7 Planning the evaluation **13**
7.1 General 13
7.2 Determine information about the system 14
7.3 Controlling factors that influence performance 15
7.4 Test subject selection 16
7.5 Test size 17
7.5.1 General 17
7.5.2 Collecting multiple recognition transactions per test subject per system 17
7.5.3 Requirements on test size 18
7.6 Multiple tests 18
8 Data collection **19**
8.1 Avoidance of data collection errors 19
8.2 Data and details collected 19
8.3 Enrolments 20
8.3.1 Enrolment transactions 20
8.3.2 Enrolment conditions 21
8.3.3 Enrolment failures and presentation errors 21
8.4 One-to-one comparison trials 22
8.4.1 General 22
8.4.2 Collection conditions 22
8.4.3 Frequency of use 22
8.4.4 Systems performing optimization based on enrolled references 23
8.4.5 Systems performing reference adaptation 23
8.4.6 Processes for data entry errors and system misuse 23

8.4.7 Failures to acquire 23
8.4.8 Adding test data to the corpus 23
8.4.9 Online comparison trials 23
8.4.10 Offline comparison trials 24
8.4.11 Offline non-mated comparison trials when references are dependent 25
8.4.12 Offline non-mated comparison trials based on comparison of references 25
8.4.13 Use of samples from multi-capture comparison transactions 25
8.5 Identification trials 26
8.5.1 General 26
8.5.2 Identification testing with non-enrolled test subjects 26
8.5.3 Use of jack-knife approach for identification testing 26

9 Analyses 26
9.1 General 26
9.2 Performance of biometric enrolment 27
9.2.1 Failure-to-enrol rate 27
9.2.2 Enrolment transaction duration 27
9.3 Performance of biometric acquisition 28
9.3.1 Failure-to-acquire rate 28
9.3.2 Acquisition process duration 28
9.3.3 Other aspects of acquisition performance 28
9.4 One-to-one comparison performance 29
9.4.1 False non-match rate 29
9.4.2 False match rate 29
9.5 Verification system performance metrics 30
9.5.1 General 30
9.5.2 False reject rate 30
9.5.3 False accept rate 31
9.5.4 Verification transaction duration 31
9.5.5 Generalized false reject rate and generalized false accept rate 31
9.6 Identification system performance metrics 32
9.6.1 General 32
9.6.2 False-negative identification rate 33
9.6.3 False-positive identification rate 33
9.6.4 Generalized false-negative identification rate and generalized false-positive identification rate 34
9.6.5 Selectivity 34
9.6.6 Closed-set test of identification performance 35
9.6.7 Estimation of identification error rates from one-to-one comparison results 35
9.6.8 Predicting identification error rates in larger populations 36
9.7 Analysis of performance across controlled experimental factors 36
9.7.1 Longitudinal analyses 36
9.7.2 Pairwise analyses 36
9.8 Detection error trade-off 36
9.9 Transaction durations 37
9.10 Computational workload 37
9.11 Uncertainty of estimates 38

10 Graphical presentation of results 39
10.1 Score distributions 39
10.1.1 General 39
10.1.2 Boxplots 39
10.2 Error rate vs threshold plot 40
10.3 DET plot 40
10.4 CMC plot / FNIR over rank plot 42
10.5 FNIR over number of enrolees plot 44
10.6 Heat maps 45

11 Record keeping 46

12 Reporting performance results **47**
12.1 Reporting test details 47
12.2 Summary statistics 48
12.3 Reporting enrolment performance 48
12.4 Reporting acquisition performance 48
12.5 Reporting one-to-one comparison performance 49
12.6 Reporting verification system performance 49
12.7 Reporting identification system performance 49
12.8 Reporting performance across factors 50
Annex A (informative) **Differences between evaluation types** **52**
Annex B (Informative) **Test size and random uncertainty** **53**
Annex C (informative) **Factors influencing performance** **61**
Annex D (Informative) **Pre-selection algorithm performance** **66**
Annex E (informative) **Identification performance as a function of database size** **68**
Annex F (informative) **Algorithms for generating DET and CMC** **70**
Annex G (informative) **DET properties and interpretation** **73**
Bibliography **77**

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.

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) or the IEC list of patent declarations received (see patents.iec.ch).

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

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html.

This document was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 37, *Biometrics*.

This second edition cancels and replaces the first edition (ISO/IEC 19795-1:2006), which has been technically revised.

The main changes compared to the previous edition are as follows:

— Terminology is updated to follow the biometrics vocabulary of ISO/IEC 2382-37:2017;

— Additional detail is provided on testing and reporting of transaction times and computational workload, and on graphical representation of results.

A list of all parts in the ISO 19795 series can be found on the ISO website.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

This document is concerned solely with the scientific technical performance testing of biometric systems and devices. Technical performance testing seeks to determine error and throughput rates, with the goal of understanding and predicting the real-world error and throughput performance of biometric systems. The error rates include both false-positive and false-negative rates, as well as failure-to-enrol and failure-to-acquire rates across the test population. Throughput rates refer to the number of individuals processed per unit of time based both on computational speed and human–machine interaction. These measures are generally applicable to all biometric systems and devices. Technical performance tests that are modality-specific, for example, fingerprint scanner image quality, are not considered in this document.

The purpose of this document is to present the requirements and best scientific practices for conducting and reporting technical performance testing. It is acknowledged that technical performance testing is only one form of biometric testing. Other types of testing not considered in this document include:

— reliability, availability and maintainability;

— security, including vulnerability;

— conformance;

— safety;

— human factors, including user acceptance;

— cost/benefit;

— privacy regulation conformance.

Biometric technical performance testing can be of three types: technology, scenario and operational evaluation. Each type of test requires a different protocol and produces different types of results. Other parts of the ISO/IEC 19795 series provide specific advice and requirements for the development and use of such different test protocols. This document addresses specific philosophies and principles that can be applied over a broad range of test conditions.

Information technology — Biometric performance testing and reporting —

Part 1: Principles and framework

1 Scope

This document:

— establishes general principles for testing the performance of biometrics systems in terms of error rates and throughput rates for purposes including measurement of performance, prediction of performance, comparison of performance, and verifying conformance with specified performance requirements;

— specifies performance metrics for biometric systems;

— specifies requirements on the recording of test data and reporting of test results; and

— specifies requirements on test protocols in order to:

  — reduce bias due to inappropriate data collection or analytic procedures;

  — help achieve the best estimate of field performance for the expended effort;

  — improve understanding of the limits of applicability of the test results.

This document is applicable to empirical performance testing of biometric systems and algorithms through analysis of the comparison scores and decisions output by the system, without requiring detailed knowledge of the system's algorithms or of the underlying distribution of biometric characteristics in the population of interest.

Not within the scope of this document is the measurement of error and throughput rates for people deliberately trying to subvert the intended operation of the biometric system (e.g. by presentation attacks).

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 2382-37, *Information technology — Vocabulary — Part 37: Biometrics*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 2382-37 and the following apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at http://www.electropedia.org/

3.1
test subject
individual whose biometric data is intended to be enrolled or compared as part of the evaluation

3.2
test crew
set of *test subjects* (3.1) utilized in an evaluation

3.3
target population
set of biometric data subjects of the application for which performance is being evaluated

3.4
test organization
functional entity under whose auspices the test is conducted

3.5
experimenter
individual responsible for defining, designing and analysing the test

3.6
test administrator
individual performing the testing

EXAMPLE Staff conducting enrolments or overseeing verification or *identification transactions* (3.10).

3.7
test observer
individual recording test data or monitoring the *test crew* (3.2)

3.8
enrolment attempt
sequence of one or more capture attempts with the aim of producing a biometric reference for a capture subject

Note 1 to entry: An enrolment attempt can require a specific number of capture attempts (e.g. three separate placements of a finger on a sensor within a set period), from which the highest quality sample(s) is/are selected for further processing.

3.9
enrolment transaction
one or more *enrolment attempts* (3.8) with the aim of producing a biometric reference for a capture subject

Note 1 to entry: If an enrolment attempt fails, further enrolment attempts can be performed within the same enrolment transaction until an attempt succeeds or enrolment is given up.

3.10
identification transaction
sequence of one or more capture attempts and biometric searches to find and return the biometric reference identifier(s) attributable to a single individual

3.11
channel effect
variation of the biometric sample due to sampling, noise and frequency response characteristics of the sensor and transmission channel

3.12
presentation effect
variation of the biometric sample due to the way that biometric characteristics are presented to the sensor

EXAMPLE In facial recognition, this can include pose angle; in fingerprinting, finger rotation and skin moisture. In many cases, the distinction between changes in the fundamental biometric characteristic and the presentation effects are unclear (e.g. facial expression in facial recognition or pitch change in speaker verification systems).

3.13
technology evaluation
offline (3.17) evaluation of one or more algorithms for the same biometric modality using a pre-existing or especially-collected corpus of samples

3.14
scenario evaluation
evaluation that measures end-to-end system performance in a prototype or simulated application with a *test crew* (3.2)

3.15
operational evaluation
evaluation that measures the performance of a biometric system in a specific application environment using a specific *target population* (3.3)

3.16
online
pertaining to execution of biometric enrolment or comparison directly following the biometric acquisition process

3.17
offline
pertaining to execution of biometric enrolment or comparison of stored biometric data subsequent to and disconnected from the biometric acquisition process

Note 1 to entry: Collecting a corpus of images or signals for offline enrolment and calculation of comparison scores allows greater control over which probe and reference images are to be used in any transaction.

3.18
closed-set test
test in which the *test crew* (3.2) comprises only individuals known to have a reference in the enrolment database

Note 1 to entry: Closed-set tests are a specific type of test for showing performance of identification systems in terms of a *cumulative match characteristic plot* (3.29).

3.19
failure to acquire
failure of the biometric capture and feature extraction processes to produce biometric features suitable for biometric comparison

3.20
false reject rate
FRR
proportion of verification transactions with true biometric claims erroneously rejected

3.21
false accept rate
FAR
proportion of verification transactions with false biometric claims erroneously accepted

3.22
false-negative identification rate
FNIR
FNIR(*N*, *R*, *T*)
proportion of a specified set of *identification transactions* (3.10) by capture subjects enrolled in the system for which the subject's correct reference identifier is not among those returned

Note 1 to entry: The false-negative identification rate can be expressed as a function of *N*, the number of enrolees, and of parameters of the identification process where only candidates up to *rank* (3.24) *R*, and with a candidate score greater than threshold *T* are returned to the candidate list.

3.23
false-positive identification rate
FPIR
FPIR(*N*, *T*)
proportion of *identification transactions* (3.10) by capture subjects not enrolled in the system for which a reference identifier is returned

Note 1 to entry: The false-positive identification rate can be expressed as a function of *N*, the number of enrolees, and parameters of the identification process where only candidates with a candidate score greater than threshold *T* are returned to the candidate list.

Note 2 to entry: For systems that always return a fixed number of candidates without applying a threshold on scores, FPIR is not a meaningful metric.

3.24
rank
position of a candidate in a candidate list ordered by descending similarity score

3.25
true-positive identification rate
TPIR
TPIR(*N*, *R*, *T*)
proportion of *identification transactions* (3.10) by capture subjects enrolled in the system for which the subject's correct identifier is among those returned

Note 1 to entry: The true-positive identification rate can be expressed as a function of *N*, the number of enrolees, and of parameters of the identification process where only candidates up to *rank* (3.24) *R*, and with a candidate score greater than threshold *T* are returned to the candidate list.

Note 2 to entry: TPIR(*N*, *R*, *T*) = 1 – FPIR(*N*, *R*, *T*).

3.26
selectivity
SEL(*N*, *R*, *T*)
average number of candidates returned above threshold *T* in a non-mated *identification transaction* (3.10)

Note 1 to entry: Selectivity can be expressed as a function of *N*, the number of enrolees, and of parameters of the identification process where only candidates up to *rank* (3.24) *R* and with candidate score greater than threshold *T* are returned on the candidate list.

Note 2 to entry: When *R* = *N*, SEL(*N*, *R*, *T*) is measured against the entire database.

3.27
computational workload
total computational effort of a single transaction (or set of transactions) in a biometric system, including number of intrinsic operations, execution time and memory requirements

Note 1 to entry: Computational workload is dependent on the hardware on which the biometric system is operating.

3.28
detection error trade-off
DET
relationship between false-negative and false-positive errors of a binary classification system as the discrimination threshold varies

Note 1 to entry: The DET can be represented as a DET table or as a DET plot.

Note 2 to entry: The receiver operating characteristic (ROC) curve was used in the previous edition of this document. The ROC is unified with the DET.

3.29
cumulative match characteristic plot
CMC plot
graphical presentation of results of mated searches a closed-set identification test, plotting the *true-positive identification rate* (3.25), TPIR(*N, R,* 0), as a function of *R*

3.30
pre-selection algorithm
algorithm to reduce the number of comparisons that need to be made in an identification search of the enrolment database

3.31
pre-selection error
<pre-selection algorithm> error that occurs when the corresponding subject identifier is not in the pre-selected subset of candidates

Note 1 to entry: In binning pre-selection, pre-selection errors occur when the data subject's enrolment reference and a subsequent sample from the same biometric characteristic are placed in different partitions.

3.32
penetration rate

<pre-selection algorithm> average proportion of the total number of references that are pre-selected

4 Abbreviated terms

API	application programming interface
CMC	cumulative match characteristic
FTAR	failure-to-acquire rate
FTCR	failure-to-capture rate
FTER	failure-to-enrol rate
FTXR	failure-to-extract rate
GFAR	generalized false accept rate
GFRR	generalized false reject rate
PIN	personal identification number
ROC	receiving operating characteristic
SDK	software developer's kit

5 Conformance

To conform to this document, a biometric performance test shall be planned, executed and reported in accordance the requirements contained in Clauses 7 through 12.

6 General biometric system

6.1 Conceptual representation of general biometric system

Given the variety of applications and technologies, it can seem difficult to draw any generalizations about biometric systems. All such systems, however, have many elements in common. Captured biometric samples are acquired from a subject by a biometric capture device and are sent to a processor that extracts the distinctive but repeatable measures of each sample (the biometric features), discarding all other components. The resulting features may be stored in the biometric enrolment database as a biometric reference. In other cases, the sample itself (without feature extraction) may be stored as the reference. A subsequent query or probe biometric sample can be compared to a specific reference, to many references, or to all references already in the database to determine if there is a match. A decision regarding the biometric claim is made based upon the similarities or dissimilarities between the features of the biometric probe and those of the reference or references compared.

Figure 1 illustrates the information flow within a general biometric system consisting of data capture, signal processing, data storage, comparison and decision subsystems. This diagram illustrates both enrolment and the operation of verification and identification systems.

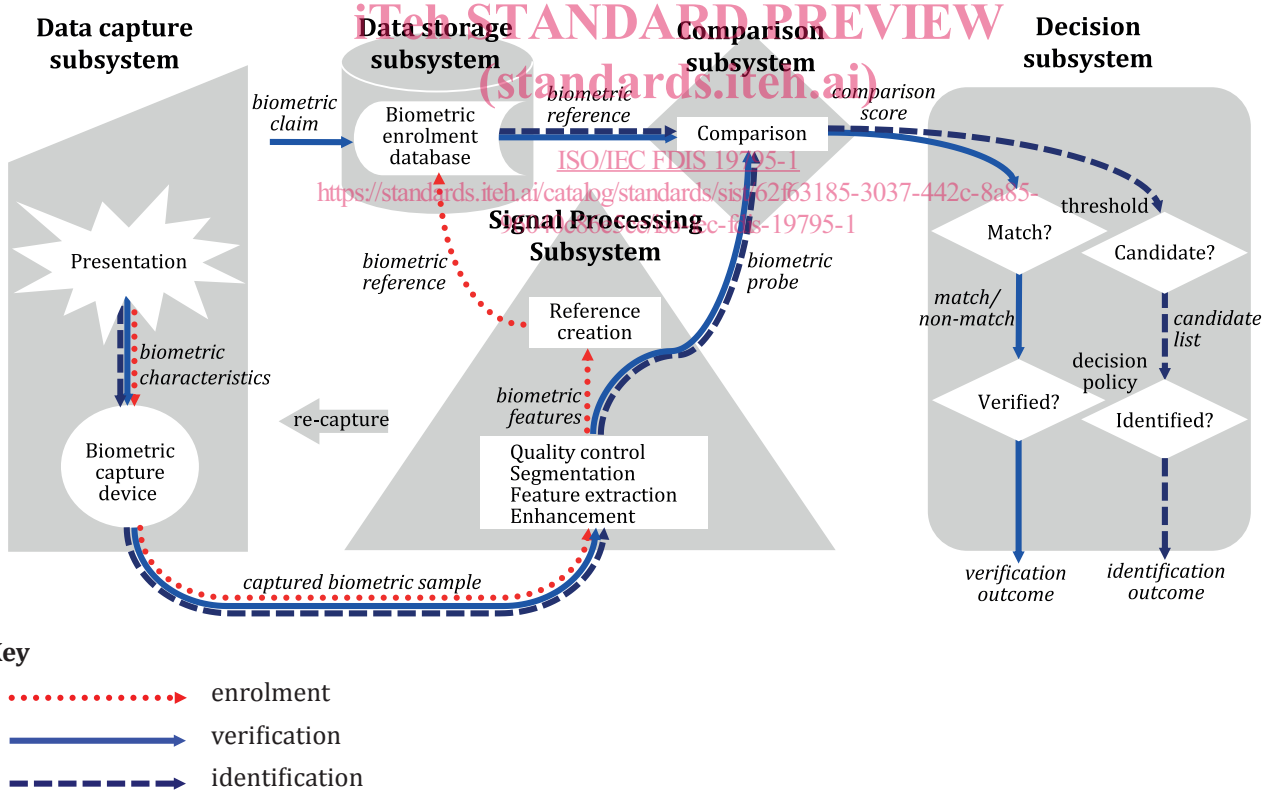

Figure 1 — Components of a general biometric system

The following subclauses describe each of these subsystems in more detail. However, it should be noted that in any implemented system, some of these conceptual components may be absent, or may not have a direct correspondence with a physical or software entity.

6.2 Conceptual components of a general biometric system

6.2.1 Data capture subsystem

The data capture subsystem collects an image or signal of a subject's biometric characteristics presented to the biometric capture device, and outputs this image or signal as a captured biometric sample.

6.2.2 Transmission subsystem

The transmission subsystem (not always present or visibly present in a biometric system) transmits samples, features, probes, references, comparison scores and outcomes between different subsystems. The captured biometric sample may be compressed and/or encrypted before transmission and expanded and/or decrypted before use. A captured biometric sample may be altered in transmission due to noise in the transmission channel as well as losses in the compression/expansion process. Data may be transmitted using standard biometric data interchange formats, and cryptographic techniques may be used to protect the authenticity, integrity, and confidentiality of stored and transmitted biometric data.

NOTE Transmission subsystem is not portrayed in Figure 1.

6.2.3 Signal processing subsystem

Signal processing includes processes such as:

— enhancement, i.e. improving the quality and clarity of the captured biometric sample;

— segmentation, i.e. locating the signal of the subject's biometric characteristics within the captured biometric sample;

— feature extraction, i.e. deriving the subject's repeatable and distinctive measures from the captured biometric sample; and

— quality control, i.e. assessing the suitability of samples, features, references, etc. and possibly affecting other processes, such as returning control to the data capture subsystem to collect further samples (recapture), or modifying parameters for segmentation, feature extraction, or comparison.

In the case of enrolment, the signal processing subsystem creates a biometric reference. Sometimes the enrolment process requires features from several presentations of the individual's biometric characteristics. Sometimes the reference comprises just the features, in which case the reference may be called a "template". Sometimes the reference comprises just the sample, in which case feature extraction from the reference occurs immediately before comparison.

In the case of verification and identification, the signal processing subsystem creates a biometric probe.

Sequencing and iteration of the above-mentioned processes are determined by the specifics of each system.

6.2.4 Data storage subsystem

References are stored within an enrolment database held in the data storage subsystem. Each reference may be associated with some details of the enrolled subject or the enrolment process. It should be noted that prior to being stored in the enrolment database, references may be reformatted into a biometric data interchange format. References may be stored within a biometric capture device, on a portable medium such as a smart card, locally such as on a personal computer or local server, in a central database, or in the 'cloud'.

6.2.5 Comparison subsystem

In the comparison subsystem, probes are compared against one or more references and comparison scores are passed to the decision subsystem. The comparison scores indicate the similarities or