BS ISO/IEC 19795-1:2021

BSI Standards Publication

Information technology — Biometric performance testing and reporting

Part 1: Principles and framework

Documents
Document Sharing Hub

Full document details: BS ISO/IEC 19795-1:2021

National foreword

This British Standard is the UK implementation of ISO/IEC 19795-1:2021. It supersedes BS ISO/IEC 19795-1:2006, which is withdrawn.

The UK participation in its preparation was entrusted to Technical Committee IST/44/-/5, Biometric testing and reporting.

A list of organizations represented on this committee can be obtained on request to its committee manager.

Contractual and legal considerations

This publication has been prepared in good faith, however no representation, warranty, assurance or undertaking (express or implied) is or will be made, and no responsibility or liability is or will be accepted by BSI in relation to the adequacy, accuracy, completeness or reasonableness of this publication. All and any such responsibility and liability is expressly disclaimed to the full extent permitted by the law.

This publication is provided as is, and is to be used at the recipient's own risk.

The recipient is advised to consider seeking professional guidance with respect to its use of this publication.

This publication is not intended to constitute a contract. Users are responsible for its correct application.

© The British Standards Institution 2021
Published by BSI Standards Limited 2021

ISBN 978 0 539 02804 1

ICS 35.240.15

Compliance with a British Standard cannot confer immunity from legal obligations.

This British Standard was published under the authority of the Standards Policy and Strategy Committee on 31 May 2021.

Amendments/corrigenda issued since publication

Date	Text affected

Documents
Document Sharing Hub

Full document details: BS ISO/IEC 19795-1:2021

INTERNATIONAL STANDARD

ISO/IEC 19795-1

Second edition
2021-05-27

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 19795-1:2021(E)

Documents
Document Sharing Hub

© ISO/IEC 2021

Full document details: BS ISO/IEC 19795-1:2021

COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2021, Published in Switzerland

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
Ch. de Blandonnet 8 • CP 401
CH-1214 Vernier, Geneva, Switzerland
Tel. +41 22 749 01 11
Fax +41 22 749 09 47
copyright@iso.org
www.iso.org

© ISO/IEC 2021 – All rights reserved

Documents
Document Sharing Hub

Full document details: BS ISO/IEC 19795-1:2021

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 ... 8
6.2.5 Comparison subsystem ... 8
6.2.6 Decision subsystem ... 8
6.2.7 Administration subsystem ... 9
6.2.8 Interface to external application ... 9
6.3 Functions of general biometric system ... 9
6.3.1 Enrolment ... 9
6.3.2 Verification of a positive biometric claim ... 10
6.3.3 Identification ... 11
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

AVA Documents
Document Sharing Hub

© ISO/IEC 2021 – All rights reserved

Full document details: BS ISO/IEC 19795-1:2021

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 ... 34
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 ... 35
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 ... 39
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

Documents
Document Sharing Hub

© ISO/IEC 2021 – All rights reserved

Full document details: BS ISO/IEC 19795-1:2021

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** **69**
Annex G (informative) **DET properties and interpretation** **72**
Bibliography **76**

© ISO/IEC 2021 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO/IEC 19795-1:2021

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 or www.iec.ch/members_experts/refdocs).

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. In the IEC, see www.iec.ch/understanding-standards.

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 and www.iec.ch/national-committees.

© ISO/IEC 2021 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO/IEC 19795-1:2021

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.

© ISO/IEC 2021 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO/IEC 19795-1:2021

This page deliberately left blank

Documents
Document Sharing Hub

Full document details: BS ISO/IEC 19795-1:2021

INTERNATIONAL STANDARD **ISO/IEC 19795-1:2021(E)**

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/

© ISO/IEC 2021 – All rights reserved

Documents
Document Sharing Hub

Full document details: BS ISO/IEC 19795-1:2021

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

© ISO/IEC 2021 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO/IEC 19795-1:2021

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

© ISO/IEC 2021 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO/IEC 19795-1:2021

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.

© ISO/IEC 2021 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO/IEC 19795-1:2021

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 in 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
FAR	false accept rate
FTAR	failure-to-acquire rate
FTCR	failure-to-capture rate
FTER	failure-to-enrol rate
FTXR	failure-to-extract rate
FNIR	false-negative identification rate
FPIR	false-positive identification rate
FRR	false reject rate
GFAR	generalized false accept rate
GFRR	generalized false reject rate

© ISO/IEC 2021 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO/IEC 19795-1:2021

PIN	personal identification number
RFID	radio frequency identification
ROC	receiving operating characteristic
SDK	software developer's kit
SEL	selectivity
TPIR	true-positive identification rate

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.

© ISO/IEC 2021 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO/IEC 19795-1:2021

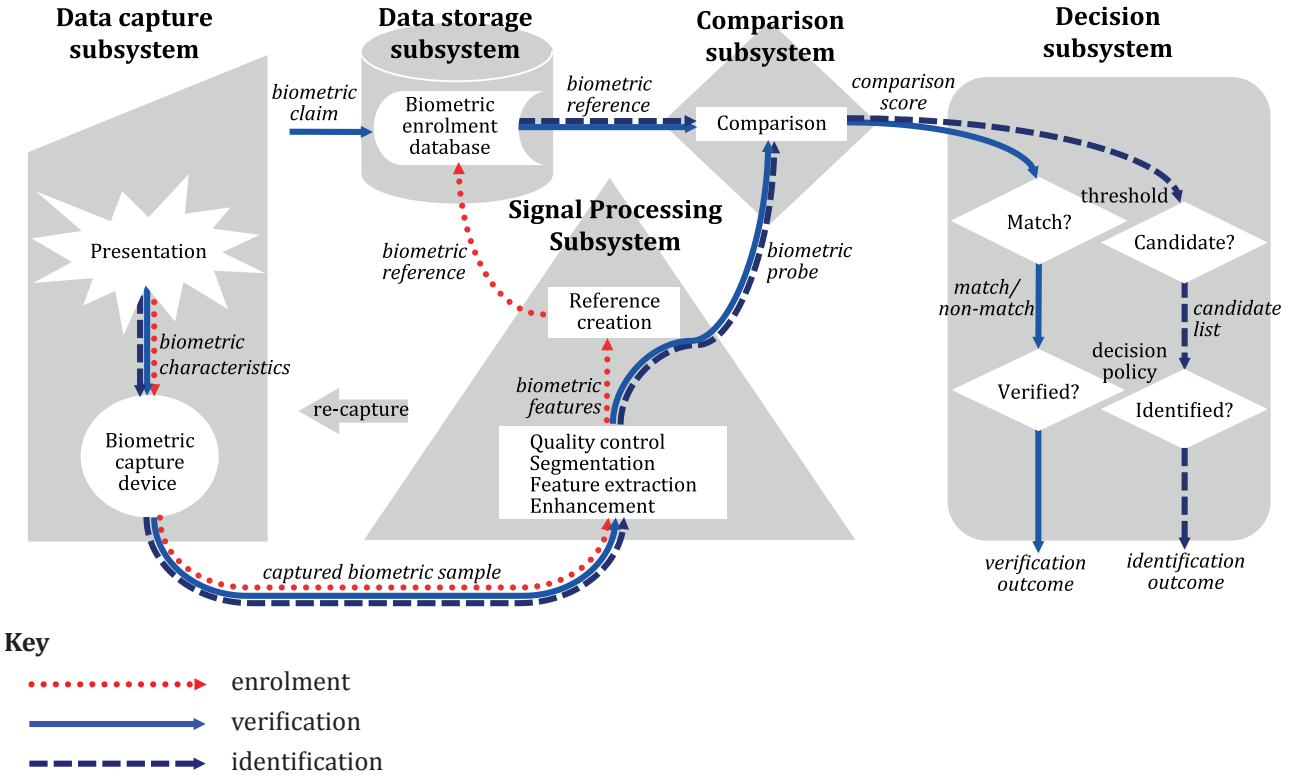

Figure 1 — Components of a general biometric system

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 The 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;

© ISO/IEC 2021 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO/IEC 19795-1:2021

— 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 dissimilarities between the probe(s) and reference(s) compared. For verification, a single specific biometric claim would lead to a single comparison score. For identification, many or all references may be compared with the probes and output a comparison score for each comparison.

6.2.6 Decision subsystem

The decision subsystem uses the comparison scores generated from one or more biometric comparisons to provide the decision outcome for a verification or identification transaction.

In the case of verification, the probes are considered to match a compared reference when (assuming that higher scores correspond to greater similarity) the comparison score exceeds a specified threshold. A biometric claim can then be verified on the basis of the decision policy, which may allow or require multiple attempts.

In the case of identification, the enrolee reference is a potential candidate for the subject when (assuming that higher scores correspond to greater similarity) the comparison score exceeds a specified threshold, and/or when the comparison score is among the predetermined number of ranked values generated during comparisons across the entire database. The decision policy may allow or require multiple attempts before making an identification decision.

NOTE Conceptually, it is possible to treat multibiometric systems in the same manner as unibiometric systems, by treating the combined captured biometric samples, references or scores as if they were a single sample, reference or score and allowing the decision subsystem to operate score fusion or decision fusion as and if appropriate. (See also ISO/IEC TR 24722.)

© ISO/IEC 2021 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO/IEC 19795-1:2021

6.2.7 Administration subsystem

The administration subsystem governs the overall policy, implementation, configuration and operation of the biometric system. Illustrative examples include:

a) interacting with the subject including providing guidance feedback to the subject during and/or after data capture, and requesting additional information from the subject;

b) storing and formatting of the biometric references and/or biometric interchange data;

c) providing final arbitration on output from decision and/or scores;

d) setting threshold values;

e) setting biometric system acquisition settings;

f) controlling the operational environment and non-biometric data storage;

g) providing appropriate safeguards for subject privacy and subject data security; and

h) interacting with the application that utilizes the biometric system.

NOTE The administration subsystem is not portrayed in Figure 1.

6.2.8 Interface to external application

The biometric system may or may not interface to an external application or system via a web services interface, an API, a hardware interface or a protocol interface.

NOTE The interface to external application is not portrayed in Figure 1.

6.3 Functions of general biometric system

6.3.1 Enrolment

In enrolment, a transaction by a capture subject is processed by the system in order to generate and store an enrolment reference for that individual.

Enrolment typically involves:

a) sample capture;

b) sample optimization or enhancement;

c) segmentation;

d) feature extraction;

e) quality checks (which may reject the sample/features as being unsuitable for creating a reference, and require capture of further samples);

f) presentation attack detection checks (which may reject the sample/features as being ineligible for use as an enrolment reference);

g) (where system policy so requires) comparison against existing biometric references to ensure the subject is not already enrolled;

h) reference creation (which may require features from multiple samples) and possibly generation of a database index;

i) storage of the biometric reference data record, possibly after conversion to a biometric reference data interchange format;

© ISO/IEC 2021 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO/IEC 19795-1:2021

j) test verification or identification attempts by the capture subject to ensure that the resulting biometric reference is usable; and

k) allowing repeat enrolment attempts, should the initial enrolment be deemed unsatisfactory (dependent on the enrolment policy).

6.3.2 Verification of a positive biometric claim

In applications such as access control, a transaction by a subject may be processed by the system in order to verify a positive specific claim about the subject's enrolment (e.g. "I am enrolled as subject X"). Note that some biometric systems allow a single subject to enrol more than one instance of a biometric characteristic (for example, an iris system may allow subjects to enrol both iris images, while a fingerprint system may require enrolment of additional fingers for fallback in case a primary finger is damaged).

Verification of a specific positive claim typically involves:

a) sample capture;

b) sample optimization or enhancement;

c) segmentation;

d) feature extraction;

e) quality checks (which may reject the sample/features as being unsuitable for comparison, and require capture of further samples);

f) presentation attack detection checks (which may reject the sample/features as being ineligible for use)

g) probe creation (which may require features from multiple samples), possible conversion into a biometric data interchange format;

h) comparison of the probe and the reference for a biometric claim producing a comparison score;

i) determination of whether the biometric features of the probe match those of the reference based on whether the comparison score exceeds a threshold (in cases where higher scores correspond to greater similarity); and

j) decision to verify a claim based on the comparison result of one or more attempts as dictated by the decision policy.

The verification function either accepts or rejects the specific positive claim. The verification decision outcome is considered to be erroneous if either a false claim is accepted (false accept) or a true claim is rejected (false reject). In this application, a false acceptance occurs if the submitted sample is wrongly matched to a stored reference not created by the data subject. A false rejection occurs if the submitted sample is not matched to a reference actually created by the data subject.

NOTE Verification of an unspecific positive claim is also quite possible with a biometric system. Such applications have been called "PIN-less verification" because no PIN or other identifier was necessary to establish that the data subject was indeed enrolled in the database. The process is as above through steps a) – g). However, steps h) – j) are somewhat different when the claim is unspecific:

h) comparison of the probe against all the references producing a score for each comparison;

i) determination of whether the biometric features of the probe match those of any reference based on whether the comparison score exceeds a threshold (in cases where higher scores correspond to greater similarity); and

j) decision to verify a claim based on the comparison results of one or more attempts as dictated by the decision policy.

© ISO/IEC 2021 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO/IEC 19795-1:2021

6.3.3 Identification

In identification, biometric samples from a capture subject are processed to generate a probe, and the enrolment database is searched to return identifiers of references similar to that probe. Identification provides a candidate list of identifiers containing zero, one, or more identifiers. Identification is considered correct when the subject is enrolled and an identifier for their enrolment is in the candidate list. The identification is considered to be erroneous if either an enrolled subject's identifier is not in the resulting candidate list (false-negative identification error), or if a transaction by a non-enrolled subject produces a non-empty candidate list (false-positive identification error).

Identification typically involves:

a) sample capture;

b) sample optimization or enhancement;

c) segmentation;

d) feature extraction;

e) quality checks (which may reject the sample/features as being unsuitable for comparison, and require capture of further samples);

f) presentation attack detection checks (which may reject the sample/features as being ineligible for use);

g) probe creation (which may require features from multiple samples) and possible conversion into a biometric data interchange format);

h) comparison against some or all references in the enrolment database, producing a score for each comparison;

i) determination of whether each compared reference is a potential candidate identifier for the capture subject, based on whether the comparison score exceeds a threshold and/or is among the highest ranked scores returned, producing a candidate list (higher scores correspond to greater similarity); and

j) an identification decision based on the candidate lists from one or more attempts, as dictated by the decision policy.

6.4 Enrolment, verification and identification transactions

The live acquisition processes for enrolment, verification or identification transactions consist of one or more capture attempts as allowed or required by the corresponding decision policy. Each capture attempt may consist of one or more presentations dependent on sensor operation, policy on sample quality, and any settings limiting the number of presentations or time permitted per attempt.

EXAMPLE 1 When a decision policy allows three attempts to verify, a transaction consists of one attempt, or two attempts if the first attempt is rejected, or three attempts if the first two attempts are rejected.

EXAMPLE 2 The enrolment process often requires the enrolee's biometric characteristics to be presented multiple times.

EXAMPLE 3 Some verification systems process a sequence of samples in a single attempt, for example: (a) collecting samples over a fixed period to find the best matching sample; or (b) collecting samples until either a match is obtained or the system times out.

EXAMPLE 4 Some biometric identification systems capture multiple samples from an individual and consolidate the results of biometric searches of each sample into a single candidate list. Typical cases are the use of multiple fingers in fingerprint identification, and the use of multiple frames in face recognition from video.

Figure 2 illustrates the relationship between presentations, attempts, and transactions.

© ISO/IEC 2021 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO/IEC 19795-1:2021

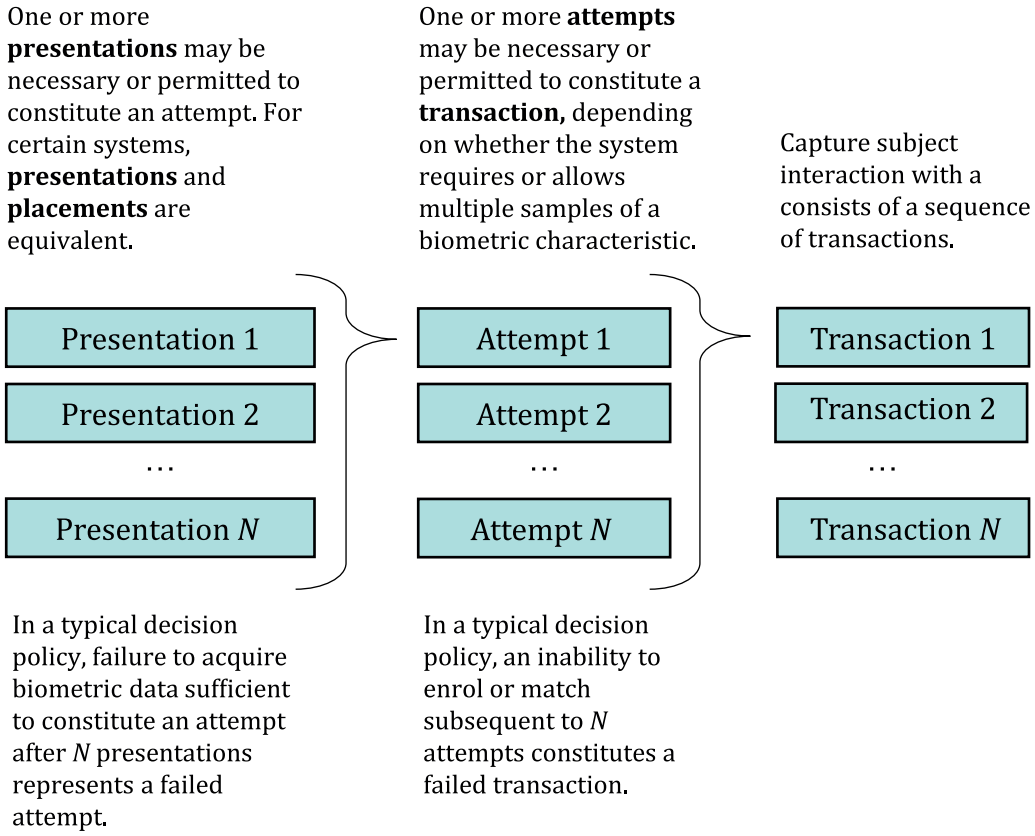

Figure 2 — Presentations, attempts and transactions

6.5 Performance measures

6.5.1 Error rates

In a biometric system, fundamental errors encompass comparison errors (false match and false non-match) and reference/probe creation errors (failure-to-capture, failure-to-extract, failure-to-enrol and failure-to-acquire). Fundamental errors combine to form decision errors for verification and identification systems. How these fundamental errors combine to form decision errors depends upon the number of comparisons required or allowed, whether there is a positive or negative biometric claim, and the decision policy.

6.5.2 Throughput rates

6.5.2.1 Throughput rates measure the number of subjects that can be processed per unit time based both on computational speed and human-machine interaction. Throughput rates for verification systems, such as those for access control, are usually determined by the speed of subject interaction with the system in the process of submitting a sufficient quality biometric sample. Throughput rates for identification systems, such as enrolment in a social service program, can also be affected by the computational processing time required to compare a captured sample to the biometric reference database. Therefore, depending on the application, it may be appropriate to measure the interaction times of subjects with the system and also the processing rate of the computational hardware as a function of the number of stored references. Further, in operational use, actions required subsequently to the biometric application decision (e.g. proceeding through a gate, opening a door, exiting a station) may affect the ability of the next subject to interact with the system. Measurement of throughput should take account of these additional actions which can vary depending on the biometric system decision.

NOTE Actual benchmark measurement of computer processing speed is covered in texts such as Reference [2] and is considered outside the scope of this document.

© ISO/IEC 2021 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO/IEC 19795-1:2021

6.5.2.2 Measurement of the speed of subject interaction requires a precise definition of the actions indicating the initiation and termination of the transaction. This definition should be determined prior to the start of the test and noted in the test report. The test report should also include a brief listing of the actions of the capture subjects included in the transaction.

EXAMPLE It is possible to measure throughput for sample acquisition only, or for recognition transactions allowing multiple attempts.

6.5.3 Types of performance testing

6.5.3.1 Biometric technical performance testing can be of three types: technology, scenario or operational.

6.5.3.2 In a technology evaluation, testing of one or more algorithms is carried out on a corpus of biometric samples. The corpus may be collected as part of a test, or it may be a previously collected corpus available to the experimenter. Although example data may be distributed for developmental or tuning purposes prior to the test, the actual testing needs to be done on data that has not previously been seen by algorithm developers. Testing is carried out using offline processing of the data. As the corpus is fixed, the results of technology tests are repeatable. Nonetheless, performance against this corpus is dependent on both the environment and the population in which it is collected.

6.5.3.3 In a scenario evaluation, testing is carried out against one or more complete systems in an environment that models a real-world target application of interest. Each tested system has its own biometric capture device and so receives slightly different data. Consequently, if multiple systems are being compared, care is required to ensure that data collection across all tested systems is in the same environment with the same population. Depending on test requirements, testing can involve offline as well as online processing of data. Test results are repeatable only to the extent that the modelled scenario can be carefully controlled.

6.5.3.4 In an operational evaluation, testing is typically online for consistency with the application environment (unless the systems under test process data in an offline fashion). Repeatability of operational test results can be limited due to unknown and undocumented differences between operational environments. Furthermore, “ground truth” on the identities and behaviours of the biometric data subjects can be difficult to ascertain, particularly under unsupervised conditions without a test administrator, test observer or operational personnel present.

Annex A summarizes the different characteristics of the different types of evaluations.

7 Planning the evaluation

7.1 General

7.1.1 As the first step in an evaluation, the experimenter shall determine:

a) the systems / application / environment / population to be evaluated;

b) the aspects of performance to be measured; and

c) how a dataset for evaluating performance is to be created or identified (i.e. which is the appropriate evaluation type: technology, scenario, or operational?).

© ISO/IEC 2021 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO/IEC 19795-1:2021

These decisions form the basis for developing an appropriate test protocol, specifying appropriate environmental controls, test subject selection and test size.

NOTE Sometimes the choice of evaluation type is determined by the availability of suitable databases of test samples for a technology evaluation, or of an installed system for an operational evaluation. Circumstances also exist in which all three types of testing are carried out sequentially, perhaps gradually narrowing down modality options and systems under consideration for the eventual deployment of a biometric system.

7.1.2 The test protocol may need to reflect system and application-specific differences such as:

a) differences in environments;

b) differences in capture subject populations (e.g. differences in subject habituation);

c) differences in biometric modalities (e.g. due to different modalities being affected by different environmental conditions, and the differences between testing of predominantly behavioural and predominantly physiological biometrics);

d) differences in the performance metrics of interest (e.g. overall performance of verification applications and identification applications are measured differently); and

e) additional problems in establishing the ground truth for identification systems (where identification transactions are not accompanied by a specific biometric claim).

7.1.3 This document provides the basic principles for conducting and reporting a performance evaluation. Other parts of the ISO/IEC 19795 series provide guidance and requirements for tests with specific evaluation types, biometric modalities and target applications.

7.2 Determine information about the system

The experimenter shall determine the following information about the system or systems to be tested in order to plan appropriate data collection procedures.

a) Does the system log transaction information? If not, then this information shall be recorded manually by the test subject, test administrator or test observer.

b) Does the system save biometric samples or features for each transaction? This is necessary if comparison scores based on data collected during the test are to be generated offline.

c) Does the system return comparison scores or just accept or reject decisions? In the latter case, data may have to be collected at a variety of threshold settings to generate a DET (see 8.2.3). If comparison scores are returned, what information is available regarding parameters and scale?

d) Is the vendor's SDK available? Offline generation of mated and non-mated comparison scores requires use of software modules from the SDK for:

   1) generating references from enrolment samples;

   2) generating probes from recognition samples; and

   3) determination of comparison scores between probes and references.

e) Are system modifications required for testing? Will required modifications alter system performance characteristics?

f) Does the system generate independent references? The correct procedures for collecting or generating non-mated comparison trials are different if references are dependent (see 8.4.11).

   NOTE An independent reference is one whose content and composition are not determined or informed by the content and composition of any other references.

© ISO/IEC 2021 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO/IEC 19795-1:2021

g) Does the system use algorithms that adapt the reference after successful verification? In such cases, consideration shall be given as to how much reference adaptation should occur prior to measuring performance and also whether non-mated comparison trials are likely to adversely affect the references (see 8.4.5).

h) What are the recommended image quality and comparison score thresholds for the target application? These settings affect the quality of presented samples and error rates.

i) Are the expected approximate error rates known? This information can help in determining whether the test size is appropriate (see B.1).

j) What are the factors that influence performance for this type of system? These shall be controlled or documented (see 7.3).

k) Does performance depend on the number of enrolled references? This is the case for most identification systems, but also for some verification systems, such as those executing cohort enrolment, or those embedding a one-to-many search within the verification process.

7.3 Controlling factors that influence performance

7.3.1 Biometric system performance figures can be highly application-, environment- and population-dependent. Annex C provides a list of population, application, environmental and system factors that have been found to affect the performance of one or more types of biometric system.

7.3.2 Factors influencing the measured performance shall be explicitly or implicitly divided into one of four classes for control:

a) factors incorporated into the structure of the experiment (as independent variables) so that their possible effects can be observed, reported and analysed;

b) factors controlled to become part of the experimental conditions (unchanging throughout the evaluation);

c) factors randomized out of the experiment; or

d) factors judged to have negligible effect, which are ignored. Without this final category, the experiment becomes unnecessarily complex.

How these factors are to be controlled shall be decided in advance of data collection. This can involve some preliminary testing of systems to determine which factors are most significant and which may be safely ignored. In determining which factors to control, there can be a conflict between the needs for internal validity (i.e. differences in performance are due only to the independent variables recorded in the study) and external validity (i.e. the results truly represent performance on the target application).

EXAMPLE Suppose we are comparing the performance of two systems and are concerned over whether the skill or personality of the test administrator affects performance. Possible ways to control this factor are: a) to design the experiment to measure the performance differential between supervisors as well as between systems; b) to use only one supervisor, or to carefully script the supervisor/subject interaction to be as consistent as possible throughout the experiment; c) to allocate enrolment attempts randomly among all supervisors, thereby avoiding any systematic bias; or d) if there is prior evidence that differences between test administrators are small compared to the differences between systems, the experiment is permitted to ignore this factor.

7.3.3 For technology testing, a generic application and population may be envisioned, ensuring that the tests are neither too hard nor too easy for the algorithms being evaluated.

7.3.4 For scenario testing, a real-world application and population should be specified and modelled in order that the biometric system can be tested on representative subjects in a realistic environment.

© ISO/IEC 2021 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO/IEC 19795-1:2021

7.3.5 In operational testing, the environment and the population are determined *in situ* with little control over them by the experimenter.

7.3.6 In scenario and operational testing, any adjustments to the devices and their environment for optimal performance (including quality and decision thresholds) should take place prior to data collection. Stricter quality control can result in fewer false matches and false non-matches, but a higher failure-to-acquire rate. The decision threshold also needs to be set appropriately if a comparison decision is presented to the capture subject — positive or negative feedback can affect subject behaviour. Vendors may advise on the optimal environment and trade-off between settings.

7.3.7 Of particular importance when planning the test is the time interval between enrolment and the collection of probe data. Longer time intervals generally make it more difficult to match probe samples to references due to the phenomenon known as "template ageing". This refers to the change in error rates caused by time-related changes in the biometric characteristic and its presentation. Collection of probe data for mated comparison trials shall therefore be separated in time from enrolment by an interval commensurate with the target application. If this interval is not known, then separation in time should be as long as is practicable.

NOTE Sometimes the elapsed time since enrolment is one of the controlled experimental factors, requiring collection of multiple samples over time. This enables a longitudinal analysis of the factors such as habituation, template ageing, and the stability/permanence of biometric characteristics over the timespan of the experiment. See 9.7.1.

7.4 Test subject selection

7.4.1 Both the enrolment and recognition functions require input signals or images. These samples should come originally from a test population or crew.

7.4.2 The test crew shall not include people whose biometric characteristics have previously been used to develop or tune the biometric system being tested.

7.4.3 The crew should be demographically similar to that of the target application for which performance will be predicted from test results. This is the case if test subjects are randomly selected from a target population. In other cases, volunteers can have to be relied upon.

7.4.4 Recruiting the crew from volunteers without imposing adequate controls can bias the tests. People with disabilities, for instance, could be under-represented in the sample population. It can be necessary to select from those volunteering to ensure that the test crew is as representative as possible and reflects the diversity of potential users. Target population approximation is an important consideration in ensuring the predictive value of tests.

7.4.5 Enrolment and recognition are normally carried out in different sessions, separated by days, weeks, months or years, depending on the target application. A test crew with stable membership over a long period is difficult to find and it should be expected that some test subjects drop out between sessions.

7.4.6 The test crew should be appropriately instructed so that their actions and behaviour follow that of the target application. If test subjects become bored or fatigued with routine testing, they can be tempted to experiment or be less careful. The experimenter shall consider test subject engagement as part of test design.

7.4.7 For applications where capture occurs without deliberate presentation, test subjects should ideally behave as if they were unaware of sample capture as it occurs. This may be achieved by passively capturing data over an extended period and by using RFID tags to establish test subjects' correct identifier without needing their input.

© ISO/IEC 2021 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO/IEC 19795-1:2021

7.4.8 Prior to testing, test subjects should be fully informed about test duration, data collection procedures, how test data is used and disseminated, and how many sessions are required. Regardless of the use of the data, the identities of the crew should never be released by the experimenter. Test subject data management shall follow local regulations about data and privacy protection. These regulations may enforce, for example, the full disclosure of information and signature of consent form before taking part in an experiment and may also enforce that all acquired data are strictly anonymized.

NOTE For some types of testing, e.g. operational testing of covert identification systems, informing test subjects is impractical or presents a risk of altering their behaviour, thereby invalidating the collected results.

7.4.9 If it is necessary to use artificially generated samples or features (including those created by modifying real data) such use shall be reported and justified, and the method of generation and assumptions regarding appropriateness shall be described. Results for synthetic and non-synthetic data shall be reported separately, and results for mixed synthetic and non-synthetic data shall report details of the mixture.

NOTE The use of artificially generated images improves the internal validity of technology evaluations, as all the independent variables affecting performance are controlled. However, external validity is likely to be reduced. The database is also likely to be biased in respect of systems that model the biometric images in a similar way to that used in their generation.

7.5 Test size

7.5.1 General

The scale of an evaluation, in terms of the number of test subjects, instances and transactions, affects how accurately error rates are measured. The larger the test, the more accurate the results are likely to be. Rules such as the Rule of 3 and Rule of 30 (described in B.1) may be used to provide lower bounds to the number of attempts or transactions needed for a given level of accuracy. However, these rules overstate the confidence associated with biometric results. They assume that error rates are due to a single source of variability, which is not generally the case with biometrics.

NOTE As the test size increases, the variance of estimates decreases, but the scaling factor depends on the source of variability. For example, subjects can have differing error rates[3], giving a component of variance that scales as 1 / (number of test subjects) instead of 1 / (number of attempts). This effect is discussed in more detail in Annex B.

7.5.2 Collecting multiple recognition transactions per test subject per system

7.5.2.1 The evaluation may collect multiple transactions from each test subject per system. Circumstances in which several transactions should be collected from each subject per system include:

— testing the effects of ageing, habituation, and other systematic variations;

— testing of systems using reference updating;

— testing the extent to which different subjects have different individual error rates; or

— when the transaction is not fully defined prior to testing, e.g. to determine how varying the number of attempts per transaction alters performance.

NOTE If the cost and effort of obtaining and enrolling the test crew does not have to be considered, the ideal test has many test subjects, each making a single transaction This provides independence between transactions. Ten probe-reference pairs from each of 100 people is not statistically equivalent to a single probe-reference from each of 1000 people and does not deliver the same level of certainty in the results. However, for cost and time efficiency, it is significantly easier to get existing enrolees to return than to find and enrol new test subjects. Furthermore, whenever a transaction is made, collecting additional transactions at the same time requires only marginally more effort. Despite the correlations between such multiple transactions, it is often the case that using multiple transactions from fewer test subjects produces a smaller uncertainty in test results than a trial of equal cost using a single transaction from slightly more test subjects.

© ISO/IEC 2021 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO/IEC 19795-1:2021