AS ISO/IEC 19795.2—2010
ISO/IEC 19795-2:2007
ISO/IEC 19795-2:2007/Amd.1:2015
(Incorporating Amendment No. 1)

Australian STANDARD®

STANDARDS Australia

Information technology — Biometric performance testing and reporting

Part 2: Testing methodologies for technology and scenario evaluation

Documents
Document Sharing Hub

Full document details: AS ISO IEC 19795.2-2010 (2021)

AS ISO/IEC 19795.2—2010

This Australian Standard® was prepared by IT-032, Biometrics and Identification. It was approved on behalf of the Council of Standards Australia on 10 May 2010.

This Standard was published on 26 May 2010.

The following are represented on Committee IT-032:

Attorney General's Department
Australian Business Limited
Centrelink
Department of Foreign Affairs and Trade
The Biometrics Institute

This Standard was issued in draft form for comment as DR 09047.

Keeping Standards up-to-date

Ensure you have the latest versions of our publications and keep up-to-date about Amendments, Rulings, Withdrawals, and new projects by visiting:
www.standards.org.au

ISBN 978 0 7337 9592 3

Documents
Document Sharing Hub

Full document details: AS ISO IEC 19795.2-2010 (2021)

AS ISO/IEC 19795.2—2010
ISO/IEC 19795-2:2007
ISO/IEC 19795-2:2007/Amd.1:2015
(Incorporating Amendment No. 1)

Information technology — Biometric performance testing and reporting

Part 2: Testing methodologies for technology and scenario evaluation

First published as AS ISO/IEC 19795.2—2010.
Reissued incorporating Amendment No 1 (October 2021).

COPYRIGHT

© ISO 2010 — All rights reserved
© Standards Australia Limited 2010

All rights are reserved. No part of this work may be reproduced or copied in any form or by any means, electronic or mechanical, including photocopying, without the written permission of the publisher.

AVA Documents
Document Sharing Hub

Full document details: AS ISO IEC 19795.2-2010 (2021)

Preface

This Standard was prepared by the Standards Australia Committee IT-032, Biometrics and Identification.

[A1] *This Standard incorporates Amendment No. 1 (October 2021). The start and end of changes introduced by the Amendment are indicated in the text by tags including the amendment number 1.* [A1]

[A1] Amendment No. 1 to this Standard was prepared by the Standards Australia Committee IT-032, Biometrics and Identification. [A1]

The objective of this Standard is to provide requirements and recommendations on data collection, analysis and reporting specific to two primary types of evaluation—technology evaluation and scenario evaluation.

[A1] This Standard is identical with, and has been reproduced from, ISO/IEC 19795-2:2007, *Information technology — Biometric performance testing and reporting — Part 2: Testing methodologies for technology and scenario evaluation* and its Amendment No. 1 (2015) which has been added at the end of the source text. [A1]

ISO/IEC 19795 consists of the following parts, under the general title *Information technology — Biometric performance testing and reporting*:

Part 1: Principles and framework

Part 2: Testing methodologies for technology and scenario evaluation

The following parts are under preparation:

Part 3: Modality-specific testing [Technical Report]

Part 4: Performance and interoperability testing of data interchange formats

Part 5: Performance of biometric access control systems

As this Standard is reproduced from an international standard, the following applies:

(a) Its number appears on the cover and title page while the international standard number appears only on the cover.

(b) In the source text 'IEC/ISO 19795' should read 'AS ISO/IEC 19795'.

(c) A full point substitutes for a comma when referring to a decimal marker.

None of the normative references in the source document have been adopted as Australian or Australian/New Zealand Standards.

The term 'informative' is used to define the application of the annexes to which it applies. An informative annex is only for information and guidance.

© Standards Australia Limited 2010

Full document details: AS ISO IEC 19795.2-2010 (2021)

Contents

Preface ii
Foreword v
Introduction vi
1 Scope 1
2 Conformance 1
3 Normative references 1
4 Terms and definitions 1
4.1 Biometric data 2
4.2 Components of a biometric system 2
4.3 User interaction with a biometric system 2
4.4 Performance measures 3
5 Overview of technology evaluations and scenario evaluations 3
6 Technology evaluation 6
6.1 Test design 6
6.1.1 Goals 6
6.1.2 Application realism 6
6.1.3 Determination of appropriate performance measures 6
6.1.4 Implementation primacy 6
6.1.5 Policies on disclosure of information to suppliers 7
6.1.6 Non-interchangeability of identification and verification attempts 7
6.1.7 Acknowledgement of models 7
6.1.8 Sequential use 8
6.1.9 Pre-test procedures 8
6.1.10 Generic test execution sequence 8
6.2 Assembling an appropriate test corpus 8
6.2.1 General 8
6.2.2 Unique enrolment 9
6.2.3 Recurrence of data acquisition 9
6.2.4 Test Subject identification 9
6.2.5 Provision of non-biometric information 9
6.2.6 Representativeness of corpus 10
6.2.7 Untainted corpus 10
6.2.8 Retirement of corpus 10
6.2.9 Corpus validation 10
6.2.10 Corpus collection environment 11
6.2.11 Failure at source 11
6.3 Performance measurement 11
6.3.1 Enrolment 11
6.3.2 Failure to acquire 12
6.3.3 Verification metrics 12
6.3.4 Identification metrics 13
6.3.5 Generalized error rates including failure to enrol and failure to acquire 13
6.3.6 Throughput performance 14
6.4 Reporting 16
6.4.1 General 16
6.4.2 System information 16
6.4.3 Data collection processes 16
6.4.4 Disclosure 17
6.4.5 Report structure 18
7 Scenario evaluation 18
7.1 Test design 18

Documents
Document Sharing Hub
Full document details: AS ISO IEC 19795.2-2010 (2021)

7.1.1 Characteristics of simulated application 18
7.1.2 Test execution 19
7.1.3 Levels of effort and decision policies 22
7.1.4 Multiple visits and transactions 23
7.1.5 Executing genuine and impostor trials 23
7.1.6 Data collection 23
7.2 Test crew 23
7.2.1 General 23
7.2.2 Habituation 23
7.2.3 Crew composition 24
7.2.4 Test Subject management 24
7.3 Performance measurement 25
7.3.1 General 25
7.3.2 Enrolment 25
7.3.3 Failure to acquire 25
7.3.4 Verification metrics 26
7.3.5 Identification metrics 26
7.3.6 Generalized error rates including failure to enrol and failure to acquire 26
7.3.7 Interim analyses 27
7.4 Reporting 27
7.4.1 General 27
7.4.2 System information 27
7.4.3 System acquisition and implementation 28
7.4.4 Physical layout of test environment 28
7.4.5 Report structure 29

8 Other issues applicable to technology and scenario evaluations 29
8.1 Parties to a test 29
8.2 Fairness 29
8.3 Basis for inclusion of test systems 30
8.4 Use of Frequently Asked Questions 30
8.5 Legal issues 30
8.6 Release of test source code 30
8.7 Supplier comment on test report 30

Annex A (informative) **Phases and activities for primary technology test types 31**

Annex B (informative) **Relationship between presentations, attempts, and transactions 42**

Annex C (informative) **Reporting effort levels 43**

Annex D (informative) **Client-server testing 45**

Annex E (informative) **Comparing results across systems in multi-system tests 47**

Amendment 1 48

Amendment Control Sheet 63

Documents
Document Sharing Hub
Full document details: AS ISO IEC 19795.2-2010 (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. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of the joint technical committee is to prepare International Standards. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

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.

ISO/IEC 19795-2 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 37, *Biometrics*.

ISO/IEC 19795 consists of the following parts, under the general title *Information technology — Biometric performance testing and reporting*:

— *Part 1: Principles and framework*

— *Part 2: Testing methodologies for technology and scenario evaluation*

The following parts are under preparation:

— *Part 3: Modality-specific testing* [Technical Report]

— *Part 4: Performance and interoperability testing of data interchange formats*

— *Part 5: Performance of biometric access control systems*

Documents
Document Sharing Hub
© Standards Australia Limited 2010

Full document details: AS ISO IEC 19795.2-2010 (2021)

Introduction

This part of ISO/IEC 19795 addresses two specific biometric performance testing methodologies: technology and scenario evaluation. The large majority of biometric tests are of one of these two generic evaluation types. Technology evaluations evaluate enrolment and comparison algorithms by means of previously collected corpuses, while scenario evaluations evaluate sensors and algorithms by processing of samples collected from Test Subjects in real time. The former is intended for generation of large volumes of comparison scores and candidate lists indicative of the fundamental discriminating power of an algorithm. The latter is intended for measurement of performance in modeled environments, inclusive of Test Subject-system interactions.

This part of ISO/IEC 19795 builds on requirements and best practices specified in ISO/IEC 19795-1, which addresses specific philosophies and principles that can be applied over a broad range of test conditions.

This part of ISO/IEC 19795 is meant to provide biometric system developers, deployers and end users with mechanisms for design, execution and reporting of biometric performance tests in a fashion that allows meaningful benchmarking of biometric performance within and across technologies, usage scenarios and environments.

© Standards Australia Limited 2010

Documents
Document Sharing Hub

Full document details: AS ISO IEC 19795.2-2010 (2021)

Australian Standard®

Information technology — Biometric performance testing and reporting

Part 2: Testing methodologies for technology and scenario evaluation

1 Scope

This part of ISO/IEC 19795 provides requirements and recommendations on data collection, analysis and reporting specific to two primary types of evaluation: technology evaluation and scenario evaluation.

This part of ISO/IEC 19795 specifies requirements in the following areas:

— development and full description of protocols for technology and scenario evaluations;

— execution and reporting of biometric evaluations reflective of the parameters associated with biometric evaluation types.

2 Conformance

A test shall claim conformance to either the technology evaluation or scenario evaluation clauses of this part of ISO/IEC 19795.

The set of clauses to which a scenario test shall conform differs from the set of clauses to which a technology test shall conform. In addition, the set of clauses to which an identification-system test shall conform differs from the set of clauses to which a verification-system test shall conform. To conform to this part of ISO/IEC 19795, an evaluation shall conform to clauses of this part of ISO/IEC 19795 as shown in Table 1.

Table 1 — Conformance for evaluation methodologies and comparison types

Evaluation methodology	Comparison type	Required clauses
Technology or scenario	Identification or verification	Clauses 5 and 8
Technology	Identification	All of Clause 6, except 6.3.3
Technology	Verification	All of Clause 6, except 6.3.4
Scenario	Identification	All of Clause 7, except 7.3.4
Scenario	Verification	All of Clause 7, except 7.3.5

3 Normative references

The following referenced documents are indispensable for the application 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 19795-1, *Information technology — Biometric performance testing and reporting — Part 1: Principles and framework*

4 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 19795-1:2006 and the following apply.

Documents
Document Sharing Hub
© Standards Australia Limited 2010

Full document details: AS ISO IEC 19795.2-2010 (2021)

4.1 Biometric data

4.1.1
biometric reference
⟨template, model⟩ user's stored reference measure based on features extracted from enrolment samples

4.2 Components of a biometric system

4.2.1
feature extractor
apparatus that extracts features from a sample

4.2.2
biometric reference generator
apparatus that transforms a sample into a biometric reference

4.3 User interaction with a biometric system

4.3.1
acclimatization
reduction, over the course of an evaluation, in a temporal condition of a biometric characteristic that may impact the ability of a sensor to process a sample

4.3.2
effort level
number of presentations, attempts or transactions needed to successfully enrol or match in a biometric system

4.3.3
enrolment attempt
submission of one or more biometric samples for a Test Subject for the purpose of enrolment in a biometric system

Note 1 to entry: One or more enrolment attempts may be permitted or required to constitute an enrolment transaction. An enrolment attempt may comprise one or more enrolment presentations.

Note 2 to entry: See Annex B for illustration of the relationship between presentation, attempt and transaction.

4.3.4
enrolment attempt limit
maximum number of attempts, or the maximum duration, a Test Subject is permitted before an enrolment transaction is terminated

4.3.5
enrolment presentation
submission of an instance of a biometric characteristic for a Test Subject for the purpose of enrolment

Note 1 to entry: One or more enrolment presentations may be permitted or required to constitute an enrolment attempt. An enrolment presentation may or may not result in an enrolment attempt.

4.3.6
enrolment presentation limit
maximum number of presentations, or the maximum duration, a Test Subject is permitted before an enrolment attempt is terminated

4.3.7
guidance
direction provided by an Administrator to a Test Subject in the course of enrolment or recognition

Note 1 to entry: Guidance is separate from feedback provided by a biometric system or device in the course of enrolment or recognition, such as audible or visual presentation queues.

© Standards Australia Limited 2010

AVA Documents Document Sharing Hub

Full document details: AS ISO IEC 19795.2-2010 (2021)

4.3.8
habituation
degree of familiarity a Test Subject has with a device

Note 1 to entry: A Test Subject having substantial familiarity with a biometric device, such as that gained in the course of employment, is referred to as a habituated Test Subject.

4.3.9
comparison attempt
submission of one or more biometric samples for a Test Subject for the purpose of comparison in a biometric system

4.3.10
comparison attempt limit
maximum number of attempts, or the maximum duration, a Test Subject is permitted before a comparison transaction is terminated

4.3.11
comparison presentation
submission of an instance of a single biometric characteristic for a Test Subject for the purpose of comparison

Note 1 to entry: One or more comparison presentations may be permitted or required to constitute a comparison attempt. A comparison presentation may or may not result in a comparison attempt.

4.3.12
comparison presentation limit
maximum number of presentations, or the maximum duration, a Test Subject is permitted before a comparison attempt is terminated

4.4 Performance measures

4.4.1
failure at source rate
proportion of samples discarded from the corpus either manually or by use of an automated biometric system prior to use in a technology evaluation

EXAMPLE A proportion of images collected in a face data collection effort may be discarded due to lack of a face in the image.

5 Overview of technology evaluations and scenario evaluations

This standard addresses two types of evaluation methodologies: technology evaluations and scenario evaluations. A test report shall state whether it presents results from a technology evaluation, a scenario evaluation, or an evaluation that combines aspects of both technology and scenario evaluations.

Technology evaluation is the offline evaluation of one or more algorithms for the same biometric modality using a pre-existing or specially-collected corpus of samples. The utility of technology testing stems from its separation of the human-sensor acquisition interaction and the recognition process, whose benefits include the following:

— Ability to conduct full cross-comparison tests. Technology evaluation affords the possibility to use the entire testing population as claimants to the identities of all other members (i.e. impostors) and this allows estimates of false match rates to be made to on the order of one in N^2, rather than one in N.

— Ability to conduct exploratory testing. Technology evaluation can be run with no real-time output demands, and is thus well-suited to research and development. For example, the effects of algorithmic

Documents
Document Sharing Hub
© Standards Australia Limited 2010
Full document details: AS ISO IEC 19795.2-2010 (2021)

improvements, changes in run time parameters such as effort levels and configurations, or different image databases, can be measured in, essentially, a closed-loop improvement cycle.

— Ability to conduct multi-instance and multi-algorithmic testing. By using common test procedures, interfaces, and metrics, technology evaluation affords the possibility to conduct repeatable evaluations of multi-instance systems (e.g. three views of a face) and multi-algorithmic (e.g. supplier A and supplier B) performance, or any combination thereof.

— Provided the corpus contains appropriate sample data, technology testing is potentially capable of testing all modules subsequent to the human-sensor interface, including: a quality control and feedback module(s), signal processing module(s), image fusion module(s) (for multi-modal or multi-instance biometrics), feature extraction and normalization module(s), feature-level fusion module(s), comparison score computation and fusion module(s), and score normalization module(s).

— The nondeterministic aspects of the human-sensor interaction preclude true repeatability and this complicates comparative product testing. Elimination of this interaction as a factor in performance measurement allows for repeatable testing. This offline process can be repeated *ad infinitum* with little marginal cost.

— If sample data is available, performance can be measured over very large target populations, utilizing samples acquired over a period of years.

NOTE 1 Collecting a database of samples for offline enrolment and calculation of comparison scores allows greater control over which samples and attempts are to be used in any transaction.

NOTE 2 Technology evaluation will always involve data storage for later, offline processing. However, with scenario evaluations, online transactions might be simpler for the tester — the system is operating in its usual manner and storage of samples, although recommended, is not absolutely necessary.

Scenario evaluation is the online evaluation of end-to-end system performance in a prototype or simulated application. The utility of scenario testing stems from the inclusion of human-sensor acquisition interaction in conjunction with the enrolment and recognition processes, whose benefits include the following:

— Ability to gauge impact of additional attempts and transactions on system's ability to enrol and recognize Test Subjects.

— Ability to collect throughput results for enrolment and recognition trials inclusive of presentation and sample capture duration.

NOTE 3 In online evaluations, the Experimenter may decide not to retain biometric samples, reducing storage requirements and in certain cases ensuring fidelity to real-world system operations. However, retention of samples in online tests is recommended for auditing and to enable subsequent offline analysis.

NOTE 4 Testing a biometric system will involve the collection of input images or signals, which are used for biometric reference generation at enrolment and for calculation of comparison scores at later attempts. The images/signals collected can either be used immediately for an online enrolment, verification or identification attempt, or may be stored and used later for offline enrolment, verification or identification.

Information on differences between technology and scenario evaluations is presented in Table 2.

© Standards Australia Limited 2010

Documents
Document Sharing Hub

Full document details: AS ISO IEC 19795.2-2010 (2021)

Table 2 — Distinctions between technology and scenario evaluations

	Technology Evaluations	Scenario Evaluations
What is tested	Biometric component (comparison or extraction algorithm).	Biometric system.
Objective of test	Measure performance of algorithm(s) on a standardized corpus.	Measure performance of end-to-end system in simulated application.
Ground truth	Known associations between data samples and source of samples, subject to data collection errors and intersections in merged data sets.	Known associations between system decisions and independently recorded sources of presented samples, subject to data collection errors and tester failure to note unwanted Test Subject behaviour.
Test Subject behaviour controlled by Experimenter	Not applicable during testing. May be known to be controlled when biometric data recorded, otherwise considered to be uncontrolled.	Controlled (unless Test Subject behaviour is an independent variable).
Test Subject has real-time feedback of the result of attempt	No.	Yes.
Repeatability of results	Repeatable.	Quasi-repeatable (if test environment conditions and human factors variables are controlled).
Control of physical environment	May be known to be controlled when biometric data recorded, otherwise considered to be uncontrolled.	Controlled and/or recorded.
Test Subject interaction recorded	Not applicable during testing. May be recorded when biometric data recorded.	Recorded.
Typical results reported	Relative robustness of biometric components or versions of components (e.g., comparison or extraction algorithms). Determine critical performance factors.	Relative robustness of biometric systems. Determine critical performance factors. Measure simulated performance.
Typical metrics	Most error rates. Not end-to-end throughput. Good for large-scale identification system performance where difficult to assemble large test crew.	Predicted end-to-end throughput. False match rate, false non-match rate. Failure to acquire, failure to enrol. GFAR, GFRR.
Constraints	Appropriate test database, e.g., gathered with one or more sensors, the identity of which may or may not be known.	Operational, instrumented system.
Human test population	Recorded.	Real time participation.

NOTE 5 Although in some cases there may be exceptions to the entries in this table, these are the main distinctions.

Documents
Document Sharing Hub
© Standards Australia Limited 2010
Full document details: AS ISO IEC 19795.2-2010 (2021)

6 Technology evaluation

6.1 Test design

6.1.1 Goals

An evaluation shall be designed to evaluate a system's enrolment, acquisition and matching functions on the target application.

6.1.2 Application realism

If the test intends to evaluate performance within an application or concept of operations, the test shall be designed and executed so that it mimics the functional (input to output) and procedural (e.g. enrolment or verification processes) aspects of such an application or concept of operations.

EXAMPLE If several images are typically gathered to constitute an enrolment transaction in a real-world enrolment attempt, technology test design should follow a similar process.

For testing purposes, the implementations under test should, if possible, return the comparison score of each comparison attempt.

6.1.3 Determination of appropriate performance measures

Experimenters shall determine which performance measures are applicable to their evaluation, in addition to those listed at clause 6.3.

Test design shall ensure that all required metrics can be generated.

Experimenters shall determine and report on the type(s) of comparison functionality to be incorporated within the technology test. One or more of the following types of comparison shall be specified:

a) verification

b) open-set identification

c) closed-set identification

The rationale for selection of one or more types of comparison functionality within a technology test shall be reported. The comparison functionality evaluated should be applicable to the algorithm in question, such that systems designed to conduct a specific type of comparison such as watchlist identification are tested in a fashion that generates the appropriate type of result.

NOTE Formulae for error rate calculation are provided in ISO/IEC 19795-1:2006, Clause 7.

6.1.4 Implementation primacy

The test plan shall not dictate the method(s) by which the biometric recognition system implements its functions. It is the responsibility of the biometric recognition implementation to perform its functions in its own way.

NOTE The separation of what a tested biometric system does from how it does it is the fundamental construct for allowing offline testing to be done. It is primarily useful in establishing the responsibilities of tester versus supplier. The system under test should be regarded wherever possible as a black box: Its essential function is to render decisions on input samples. The internal details of how this occurs may be proprietary, but in any case, are of no concern to the tester. This construction facilitates the testing of arbitrary biometric samples.

© Standards Australia Limited 2010

Documents
Document Sharing Hub

Full document details: AS ISO IEC 19795.2-2010 (2021)

EXAMPLE 1 If a fingerprint is sampled at 1000 dpi, and a test device is known to process only half that, then the tester should a) not execute the down-sampling on the basis that the method for doing so is non-trivial, and b) apprise the supplier of the need to handle the down-sampling internally.

EXAMPLE 2 A set of simultaneously acquired non-frontal face images could be processed by a biometric system and device in at least three ways: selection of the best image; fusion of all images; or stereoscopic synthesis of a 3D model. In any case the biometric system or device decides.

EXAMPLE 3 Most automated fingerprint identification system (AFIS) machines (i.e. those identifying multi-fingerprint records) implement some binning mechanism to partition the database according to some criterion (most simply, the Henry class) and to search only that part of the database of the same category as a user or impostor sample, thereby obtaining throughput benefits, but possibly incurring accuracy losses. Such a tradeoff is achieved by the supplier setting internal binning parameters, and measured by conducting full-scale repeats of the test for each configuration.

EXAMPLE 4 In a study seeking to demonstrate the utility of multiple fingers' prints in a recognition system, the tester should not pass separate samples through the device and perform subsequent score-level fusion, but should instead compose all the imagery as a sample (e.g. in an American National Standards Institute [ANSI]-National Institute of Standards and Technology [NIST] record, or a Common Biometric Exchange Formats Framework [CBEFF] wrapped instance of 19794-2) so as to let the biometric device perform the fusion internally. See ISO/IEC 19794-2, *Information technology — Biometric data interchange formats — Part 2: Finger minutiae data* for further information on CBEFF wrapped instances. See ANSI/NIST-ITL 1-2000 NIST Special Publication 500-245 for information on ANSI-NIST records.

6.1.5 Policies on disclosure of information to suppliers

The tester shall formulate policies before testing begins that govern what information will be disclosed to the suppliers a) before test equipment is configured, shipped or installed, and b) at execution time.

6.1.6 Non-interchangeability of identification and verification attempts

Comparison scores that result from a one-to-many identification search shall not be presented as the results of verification attempts without justification.

NOTE 1 The principle of operational realism indicates that performance shall be estimated from outcomes of attempts (i.e. rejects and acceptances). A verification system shall be evaluated on the results of a sequence of user claims of identity. Likewise an identification system shall be tested over one-to-many searches. Even in the case that a one-to-many search produces a full candidate list, the candidate list is atomic, meaning that it should not be regarded as the result of N verification attempts (to be used in the computation of verification performance).

NOTE 2 The non-equivalency of a single identification attempt and N one-to-one verifications arises because verification can be improved by comparing the user sample with additional hidden samples in a process known as cohort-normalization. The method adjusts the raw single comparison score in order to drive down false accept rates by effectively setting user-specific thresholds. The method trades off performance for throughput because the additional comparisons mean 1:1 verification incurs the expense of 1:M, where M is the size of the hidden biometric reference set.

NOTE 3 The use of cohort normalization is properly conducted internally to the device, making private use of an internally selected enrolled population.

6.1.7 Acknowledgement of models

If a model, approximation or prediction of identification performance is reported in place of, or in addition to, an empirical trial, the model shall be verified to the extent possible with the available data and fully documented.

Documents
Document Sharing Hub
© Standards Australia Limited 2010
Full document details: AS ISO IEC 19795.2-2010 (2021)

6.1.8 Sequential use

The test plan shall define the order of use of the test data. This order shall be appropriate to the application. The implementation should process the test data in this sequence.

NOTE 1 Transactions are ordinarily executed separately. Therefore the implementation would need to complete one transaction before commencing the next transaction.

NOTE 2 The majority of biometric applications involve the sequential and separate use of the biometric systems or device by individuals, subsequent, in the case of genuine users, to prior enrolment.

NOTE 3 Certain identification tasks may not be sequential. For example batch identification of all persons in a closed room is easier because it reduces to the linear assignment problem.

6.1.9 Pre-test procedures

6.1.9.1 Installation and validation of correct operation

The test organization shall take steps to ensure that the hardware/software is installed and configured appropriately and shall verify that the system is operating correctly.

NOTE Installation, configuration, and verification of system operations may involve supplier(s).

6.1.9.2 Data preparation

Data preparation shall ensure that Test Subject-identifying information and any associated metadata that would not ordinarily be available to the application (e.g. sex, age) is expunged from the samples. Otherwise a supplier might deduce the true identities to game the test.

6.1.10 Generic test execution sequence

The following is a generic description of the sequence of technology test execution:

— Enrolment samples are converted to biometric references and may be stored in a linear collection.

— Identification and verification samples are converted to sample features.

— Verification attempts are a direct comparison of sample features to a biometric reference.

— Closed-set identification attempts are a search of the enrolled population intended to return the user's identifier.

— Open-set identification attempts search the enrolled database and

  — return one or more identities;

  — return a null identity, indicating Test Subject is not found in the enrolled database.

NOTE 1 The above functionality may be implemented at the API level, or by scripting around executables.

NOTE 2 Annex A describes test execution sequences for specific types of technology tests.

6.2 Assembling an appropriate test corpus

6.2.1 General

Technology evaluation is designed to evaluate one or more biometric algorithms for enrolment and comparison performance. Technology test planning is contingent on the type of data an Experimenter wishes to generate.

© Standards Australia Limited 2010

Documents
Document Sharing Hub

Full document details: AS ISO IEC 19795.2-2010 (2021)

6.2.2 Unique enrolment

All corpus samples should correspond to real people. An evaluation design should not intentionally enrol different samples from the same individual as if they were from different individuals. For tests in which each identity corresponds to a different individual, the testing organization shall report processes implemented to ensure this.

If it is possible that an individual has multiple identities in the corpus, the corpus may be "cleaned" to reconcile such instances if practical. Otherwise the test should proceed under the assumption that each identity corresponds to a different individual.

NOTE 1 Biometric systems are intended to uniquely identify single individuals. If more than one image or signal is available for an individual it should be encapsulated as a single sample and used for enrolment or comparison.

NOTE 2 Populating an identification system with more than one sample per individual (from within one or more modalities) and then regarding the enrolments as nominally separate is a deprecated practice for the following reasons:

Identification entails a search through enrolled samples and generation of a candidate list. When multiple samples are separately enrolled a score-level fusion of each user's samples using the max criterion is implied because the largest scored entry wins. Even if the number of samples per person is equal for all persons, the practice is deprecated because it is the supplier's responsibility to combine each individual's samples in what it deems the best way.

Error metrics that depend on the size of the enrolled population, N, will be incorrect if the size of the enrolled population is not the number of separate individuals.

NOTE 3 An evaluation that seeks to investigate the effect of multiple (separate) enrolled biometric references per Test Subject is exempted from this clause, provided that this is documented in both the test plan and the test report.

6.2.3 Recurrence of data acquisition

Depending on the Experimenter's level of access to the test population, each Test Subject may be able to provide data multiple times over the course of multiple visits. The number of transactions and visits can be maximized in order to enable granular measurement of biometric reference aging effects though this will also be informed by habituation effects.

6.2.4 Test Subject identification

The Experimenter shall report information related to Test Subject identification, including at a minimum the following:

a) Types of identifiers used to identify Test Subjects

b) Amount and type of personal data collected

6.2.5 Provision of non-biometric information

If available in the corpus, metadata normally available to a deployed system shall be provided to the system(s) under test. The test report shall state the names and types of any metadata variables that were made available to the systems under test.

EXAMPLE Such data may be sensor-specific (e.g. sensor settings), environmental (e.g. temperature, humidity), Test Subject-specific (e.g. gender, age), or any other germane information.

NOTE Technology testing is unable to incorporate multiple aspects of real-world biometric operations, but evaluation design should not exclude aspects of real-world biometric operations unnecessarily.

Documents
Document Sharing Hub
© Standards Australia Limited 2010
Full document details: AS ISO IEC 19795.2-2010 (2021)

6.2.6 Representativeness of corpus

Evaluation design shall consider, and a test report shall document, whether the data in the test corpus is appropriate for the goals of the test or the applications of interest.

If data is acquired under the supervision or control of the test organization, information pertaining to Experimenter-Test Subject interaction shall be recorded in the areas of acclimatization, training, habituation, and guidance.

NOTE 1 The utility of technology evaluation in producing predictive estimates of deployed performance is predicated upon the assumption that it is possible to consistently acquire samples from users in the same format and with the same quality as the data used in the test.

NOTE 2 Ideally, data collected for different modalities has equivalent levels of habituation, acclimatization, guidance, etc.

6.2.7 Untainted corpus

A corpus may be considered "tainted" to a greater or lesser extent if

a) any implementation supplier has had possession of the corpus;

b) any implementation supplier has provided equipment used in collecting or processing the corpus, particularly if this activity influenced the nature or quality of the corpus such as by excluding samples;

c) a system being tested has previously been tested and tuned using the corpus.

When use of a tainted corpus is unavoidable, this fact shall be documented in the test report.

Sample data should not be used in an evaluation if one or more of the participating suppliers has had possession of it. Previous testing / tuning of the system using the test corpus (in whole or in part) shall be documented in the test report.

NOTE 1 This clause is necessary because performance may be improved via gaming.

NOTE 2 Note that it is generally insufficient to trivially alter the sample to skirt the reuse prohibition. Gaming may still be possible if any identifiable trait of the previously seen samples remains.

6.2.8 Retirement of corpus

Samples should not be reused in an evaluation if one or more systems under test have been tuned on the basis of performance measured in a previous test with that data.

NOTE 1 This is most readily achievable by using sequestered data.

NOTE 2 This may be expensive in that it implies additional collection activity.

6.2.9 Corpus validation

Validation is the process whereby Test Subject data is screened for the purpose of removal of data not suitable for the purpose of the evaluation.

Validation may include checking to ensure that Test Subject data is present, that the data is in the correct format, that the correct instance has been collected, and that ground truth errors are identified.

Experimenters shall report whether Test Subject data has been validated. If data has been validated, the Experimenter shall detail the method(s) applied in validating the data. The proportion and criteria for data removal shall be reported.

EXAMPLE 1 Database quality control might be used to avoid images with bad contrast in the Test Subject data.

© Standards Australia Limited 2010

AVA Documents Document Sharing Hub

Full document details: AS ISO IEC 19795.2-2010 (2021)

EXAMPLE 2 Data samples might be excluded that do not show a face in face recognition technology (e.g. no face at all or full body) or that do not show a fingerprint in a fingerprint recognition technology test (e.g. a palm print).

NOTE 1 Since some types of biometric data may be more easily validated than others, use of data validation could introduce a bias in performance results.

NOTE 2 Data removed by "corpus validation" is distinct from that discarded as "Failure at source." Sometimes a judgement call will be needed as to whether excluded data should be considered invalid or failure at source.

6.2.10 Corpus collection environment

Environmental conditions present during data collection may be known or specified. Such collection would typically be intended to measure performance under specific environmental conditions relative to baseline environmental conditions. Such controls may be established for temperature, lighting, humidity, and other factors known or suspected to impact biometric performance.

Available information pertaining to environmental conditions during corpus acquisition relevant to the modalities under evaluation should be reported, such as the following:

— temperature;

— exposure to elements;

— lighting, including type, direction, intensity;

— ambient noise;

— vibration.

If applicable, Experimenters shall report that such information was not available.

NOTE See ISO/IEC 19795-1:2006, C.2.6 for information on environmental factors that can impact performance.

6.2.11 Failure at source

Offline tests use stored biometric samples, which may have been gathered with or without a biometric system in the acquisition process. The test report shall disclose any known information about how the data was processed at any stage before use in the test. Particularly if samples were discarded, either manually or by use of an automated biometric system, then a *failure at source* rate (FAS) shall be reported.

NOTE 1 FAS may relate to a different biometric sensor or image quality assessment algorithm than the system under test.

NOTE 2 A judgment call may be needed: For example if a few legacy image samples are found to be entirely blank then these could be legitimately not counted in FAS, unless of course such samples would routinely occur in the application that the test is intended to mimic.

6.3 Performance measurement

6.3.1 Enrolment

Offline tests shall record, as the failure to enrol rate (FTE), the proportion of Test Subjects for whom an implementation elects to reject enrolment of their designated enrolment samples in the corpus. Criteria by which failure to enrol is declared shall be defined.

NOTE 1 Failure to enrol measured in a technology test is only partially representative of the failure modes possible in a live acquisition.

Documents
Document Sharing Hub
© Standards Australia Limited 2010
Full document details: AS ISO IEC 19795.2-2010 (2021)

NOTE 2 A failure to enrol can be declared by a system for any reason. A frequent reason is that the system (as configured with its native image or signal detection and processing capability and with some quality acceptance criteria) fails to detect the needed signal on the basis of low quality.

NOTE 3 By declaring a failure to enrol, a system can attain better comparison performance. This trade-off must be accounted for by combining failure to enrol and false non-match rate to produce generalized false reject rate (GFRR).

The Experimenter shall specify the minimum number of samples required, and the maximum number of samples permitted, for successful enrolment.

For each biometric system tested, the Experimenter should calculate the following:

a) distribution of enrolment quality scores, if available;

b) failure to enrol for different demographic groups, or associated with different environmental conditions, or for other logical segments of the corpus.

6.3.2 Failure to acquire

Offline tests shall record the proportion of verification or identification attempts for which the system fails to capture or locate an image or signal of sufficient quality. This is the *failure to acquire* rate (FTA).

NOTE 1 The failure to acquire rate is the comparison-phase analogue of the enrolment-phase failure to enrol, and thus Notes 1 and 2 in Clause 6.3.1 apply analogously.

NOTE 2 The failure to acquire must be used along with false match rate to compute false accept rate.

NOTE 3 In a technology test, FTA is typically declared by an encoding or comparison component and is attributable to failure to process an attempt.

The Experimenter shall specify the minimum number of samples required, and the maximum number of samples permitted, to create sample features.

The formula for calculating FTA can be found in ISO/IEC 19795-1.

6.3.3 Verification metrics

For each verification system tested, the Experimenter shall compute the following:

a) false match rate (FMR) and false non-match rates (FNMR);

b) false reject rate (FRR) and false accept rates (FAR), unless test design is such that false accept rate and false reject rate are identical to false match rate and false non-match rate;

c) number of genuine and impostor comparisons executed;

d) for genuine Test Subjects, distribution of time lapsed between enrollment and acquisition of sample features, if available;

e) uncertainty of test results, as well as basis and formulae for estimating uncertainty.

False match rates and false non-match rates, as well as false accept and false reject rates, may be rendered in the form of a receiver operating characteristic (ROC) or detection error tradeoff (DET) curve. The number of Test Subjects and transactions used to arrive at these rates shall be computed.

NOTE For systems that return match/non-match decisions as opposed to comparison scores, performance may be reported at a single operating point on the ROC or DET.

For verification systems, the Experimenter should calculate the following:

a) distribution of comparison scores for genuine Test Subject and impostors;

© Standards Australia Limited 2010

AVA Documents Document Sharing Hub

Full document details: AS ISO IEC 19795.2-2010 (2021)

b) verification results for different demographic groups, or associated with different environmental conditions, or for other logical segments of the corpus.

6.3.4 Identification metrics

For all identification systems, the Experimenter shall calculate uncertainty of test results, as well as basis and formulae for estimating uncertainty.

For closed-set identification systems, the Experimenter shall calculate the following:

a) cumulative match characteristics (CMC);

b) number of searches executed.

For open-set identification systems, the Experimenter shall calculate the following:

c) false positive identification rates (FPIR) and corresponding false negative identification rates (FNIR) (preferably over a range of thresholds);

d) binning error rate and penetration rate if binning is used.

For identification systems, the Experimenter shall calculate the following:

e) identification results for different demographic groups, or associated with different environmental conditions, or for other logical segments of the corpus.

6.3.5 Generalized error rates including failure to enrol and failure to acquire

6.3.5.1 General

The immediate output of an offline test — a set of paired (false match rate, false non-match rate) values — shall be combined with the measured values of failure to acquire and failure to enrol.

NOTE 1 Because a system can improve its false acceptance and rejection performance by abstaining from processing low quality samples, it is necessary to combine failure to acquire and failure to enrol rates with the measured false match rate and false non-match rate to produce the final statement of performance.

If FTE and FTA are known to be zero this fact should be noted. In this case, GFAR and GFRR for single-attempt transactions do not differ from FMR and FNMR. If FTE or FTA are known to be non-zero, false accept rate and false reject rate should be computed, and thereby differentiated from, false match rate and false non-match rate.

NOTE 2 In certain tests, samples that resulted in failure to acquire or failure to enrol may be released to suppliers for further study.

NOTE 3 For systems that return match/non-match decisions as opposed to comparison scores, performance may be reported at a single operating point on the ROC or DET.

NOTE 4 By declaring very many failures to enrol or to acquire, an implementation under test may achieve low GFAR values, but will thereby increase GFRR.

6.3.5.2 Single-attempt transactions

For each implementation under test, the experimenter shall determine the generalised FAR (GFAR) for single-attempt transactions and the generalised FRR (GFRR) for single-attempt transactions.

Documents
Document Sharing Hub
© Standards Australia Limited 2010
Full document details: AS ISO IEC 19795.2-2010 (2021)

In cases where transactions consist of single attempts, generalised false accept rate may be computed as the proportion of impostors who are acquired and matched at some operating point threshold, t:

GFAR(t) = (1-FTA) FMR(t) (1-FTE)

Similarly the generalised false reject rate is that proportion of genuine users who are either unable to be acquired during use, or who can acquire but can't enrol, or who can enrol, be acquired and are falsely rejected, at some operating point threshold, t:

GFRR(t) = FTA + (1-FTA) FTE + (1-FTA) (1-FTE) FNMR(t)

The given formulas for GFAR and GFRR hold only in the special case of n=1, where n is the number of attempts allowed in a transaction.

NOTE 1 Different formulae may be needed if it is acceptable to not enrol certain individuals.

NOTE 2 Explicit failure to enrol and failure to acquire measurements can be avoided by specifying that all verification comparisons will result in a comparison score. A supplier can satisfy this requirement by internally recording a failure-to-enrol or failure to acquire condition and report suitably low values when such a biometric reference is used in a one-to-one comparison. This method correctly includes failure to enrol and acquire in the DET characteristic.

NOTE 3 Alternatively, GFAR and GFRR can also be determined by:

— including failed (neither accept nor reject) impostor transactions and impostor transactions for individuals whose enrolment failed in the total number of impostor transactions,

— including failed (neither accept nor reject) genuine transactions and genuine transactions for individuals whose enrolment failed in the total number of genuine transactions, and

— counting failed (neither accept nor reject) genuine transactions and genuine transactions for individuals whose enrolment failed as false rejects.

6.3.5.3 Multi-attempt transactions

In cases where transactions consist of multiple attempts, GFAR and GFRR calculations are more complex. Formulae for such tests should be generated on a test-specific basis.

6.3.6 Throughput performance

6.3.6.1 General

The testing organization may measure throughput suitable to the implementation under test.

If transaction time is an aspect of performance that a test intends to measure, then the experimenter shall specify a method of measuring transaction time suited to the implementation under test.

NOTE 1 Ideally, all enrolment and comparison operations would be timed.

NOTE 2 Offline testing will address only the computational components of throughput. For example in enrolment an offline test will only capture the image analysis and biometric reference creation phase of the full enrolment process while excluding ergonomic and transactional aspects of human related activities (e.g. placements on the sensor, removing eye glasses). Thus the enrolment throughput rates as measured in a technology test represent a lower bound on operational throughput rates.

6.3.6.2 Reporting throughput performance

If throughput summary statistics are calculated, the mean shall be reported. Other summary statistics such as the following may also be reported:

a) minimum;

© Standards Australia Limited 2010

AVA Documents Document Sharing Hub

Full document details: AS ISO IEC 19795.2-2010 (2021)

b) maximum;

c) median;

d) standard deviation.

NOTE If populations of different sizes are enrolled (particularly in identification trials) the tester should report sufficient timing information to allow an assessment of the functional dependency on population size to be made; for example O(N) or O(N^2).

6.3.6.3 Reporting comparison and throughput performance

Throughput performance is vital in technology evaluation because, in general, recognition errors can be reduced if throughput is decreased. In such cases a full statement of performance would be a DET characteristic with an additional third axis, throughput rate, allowing a deployer to select a suitable operating point. Moreover, with many biometric systems, altering the decision threshold will require that more, or fewer, presentation samples to be needed for successful recognition of a genuine attempt, thereby affecting throughput.

6.3.6.4 Measuring biometric reference generation and sample feature extraction timing

Some systems are asymmetric in that generation of biometric references uses different kinds of samples and different algorithmic processes than those used in feature extraction for verification and identification. Timings for biometric reference generation and for feature extraction may therefore be reported separately.

6.3.6.5 Simultaneous measurement of throughput and recognition error rates

Throughput measurement may be made during the same trial as the recognition error rates, and statistics may be bound to them.

6.3.6.6 Throughput in impostor and genuine user attempts

By the principle of operational realism user throughput may be measured for both impostor and genuine user attempts, and statistics may be reported separately.

6.3.6.7 Post-enrolment “fixing” overhead

In identification trials the tester should be cognizant that after a population is enrolled a post enrolment “fixing” overhead may be incurred. Fixing here means any process that a system invokes at the end of enrolment such is typically used to separate feature vectors for better performance.

6.3.6.8 Uniqueness searches on enrolment

Real-world enrolment of a population often embeds a uniqueness search for each new candidate enrolee. This implies that the enrolment of a population of size N, has O(N^2) expense. In a technology test the uniqueness of the samples is usually assured by design, such that enrolment has O(N) cost.

The experimenter should determine whether 1:N uniqueness determination is a component of the enrollment process. This might be ascertained through measurement of enrollment durations as the size of the enrolled database increases. If such behavior is observed or known to be implemented, then the experimenter should report results accordingly. To separate the time required to determine uniqueness from the fixed enrollment time, a test design may require that implementations disable duplicate detection functionality.

NOTE A dedicated enrolment test, building a database from scratch, can be conducted.

Documents
Document Sharing Hub
© Standards Australia Limited 2010
Full document details: AS ISO IEC 19795.2-2010 (2021)