INTERNATIONAL STANDARD

ISO/IEC 19795-2

First edition
2007-02-01

AMENDMENT 1
2015-04-01

Information technology — Biometric performance testing and reporting —

Part 2: Testing methodologies for technology and scenario evaluation

AMENDMENT 1: Testing of multimodal biometric implementations

Technologies de l'information — Essais et rapports de performance biométriques —

Partie 2: Méthodologies d'essai pour l'évaluation des technologies et du scénario

AMENDEMENT 1: Essais des mises en oeuvre biométriques multimodales

Reference number
ISO/IEC 19795-2:2007/Amd.1:2015(E)

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

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

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 ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

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

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

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

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

Introduction

Testing and reporting methods in this part of ISO/IEC 19795 are primarily intended for single-modality systems. These standards can be inadequate for executing reproducible performance evaluations of multimodal biometric systems such as those used in border control applications. Various configurations are proposed for multimodal biometric systems, as described in ISO/IEC TR 24722. It is necessary to clearly identify methods and requirements for multimodal biometric systems evaluation such as variation of parameters and environmental factors that are to be described when reporting.

Information technology — Biometric performance testing and reporting —

Part 2: Testing methodologies for technology and scenario evaluation

AMENDMENT 1: Testing of multimodal biometric implementations

Add the following items to the list in the scope clause of ISO/IEC 19795-2*:*

— multimodal biometric specific requirements for technology evaluation and scenario evaluation;

— description of the structure and performance measures of multimodal biometric devices and systems;

— specification of biometric data collection and performance calculation methods;

— specification of reporting requirements.

Add the following item to the definitions clause in this part of ISO/IEC 19795*:*

4.4.2

multimodal FTE

MFTE

proportion of the population for whom the multimodal biometric system with a stated enrolment policy fails to complete the enrolment process

EXAMPLE 1 For a policy which allows one of a set of modalities to be enrolled, MFTE would be the proportion of subjects that fail to enrol in all the modalities. This results in lower effective FTE.

EXAMPLE 2 For a policy which requires enrolment in all considered modalities, MFTE would be the proportion of subjects that fail to enrol in one or more modalities. This results in an effective FTE greater than or equal to the higher of the individual modality FTEs.

Insert the following normative Annex into this part of ISO/IEC 19795:

Annex F
(normative)
Testing of multimodal biometric implementations

F.1 General

This annex specifies methods for evaluating and reporting the performance of multimodal biometric algorithms and systems.

Multimodal biometric implementations might be used to meet the following objectives:

— to support users who cannot present one or more requested modalities to the system, in other words, to improve failure-to-enrol rate;

— to improve biometric system throughput;

— to improve recognition performance (e.g. through reduction of false negative identification rates);

— to improve usability; and

— to increase robustness against presentation attacks.

ISO/IEC TR 24722 defines the following multimodal fusion levels:

— decision-level;

— score-level;

— feature-level;

— sample-level.

Multimodal fusion implementations differ across each level. Even when multimodal data are gathered with identical sensors, results might differ based the fusion level implemented.

For this reason, the experimenter shall determine the system or application to be evaluated. An evaluation shall clearly identify the fusion level implemented, the components of the multimodal implementation, and requirements applicable to evaluations for each fusion level.

Two types of multimodal evaluations can be considered:

— evaluations in which the experimenter does require insight into the multimodal system;

— evaluations in which the experimenter does not require insight into the integrated multimodal system.

This annex focuses on evaluations in which the experimenter requires insight into the multimodal system. F.1 and F.6 are applicable to all multimodal implementations. If the experimenter does not require insight into the integrated multimodal system component shown in Figure F.2 or Figure F.4, then the multimodal implementation can be tested without using this annex.

F.2 Fusion scheme identification information for repeatable evaluation

F.2.1 Decision-level fusion

F.2.1.1 General

An example of decision-level fusion is shown in Figure F.1. Decision-level fusion systems combine decision results from separate biometric sub systems.

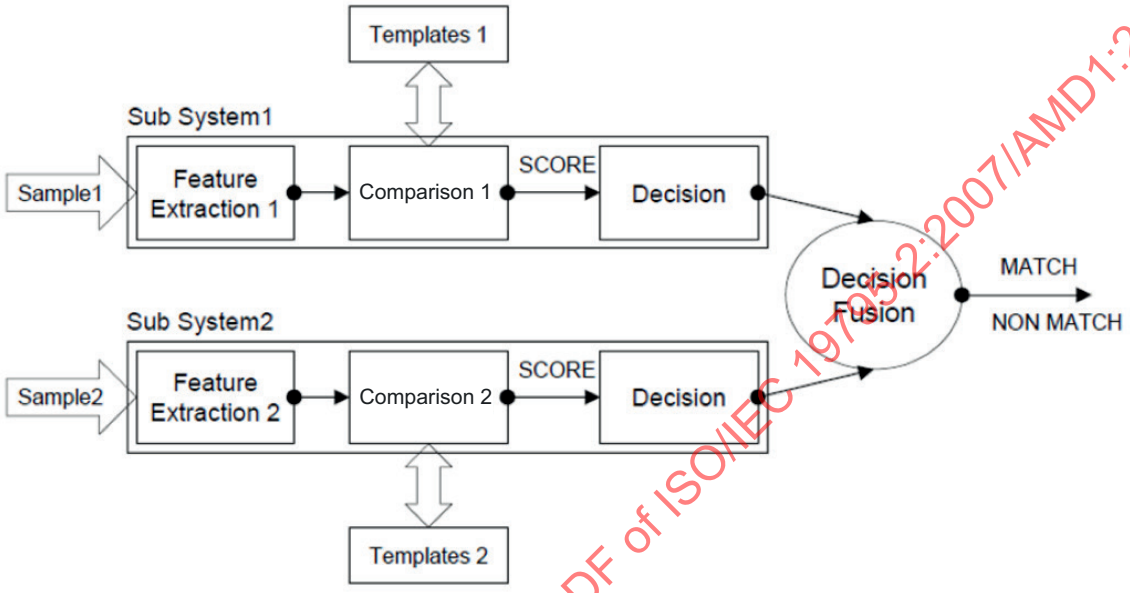

Figure F.1 — Decision-level fusion

NOTE Decision-level fusion systems might be used to improve false match rate (FMR) or false non-match rate (FNMR).

Figure F.2 is an example of an integrated multimodal system with a decision output in which insight is not provided into the implementation. This type of system can be tested without methods described in this annex.

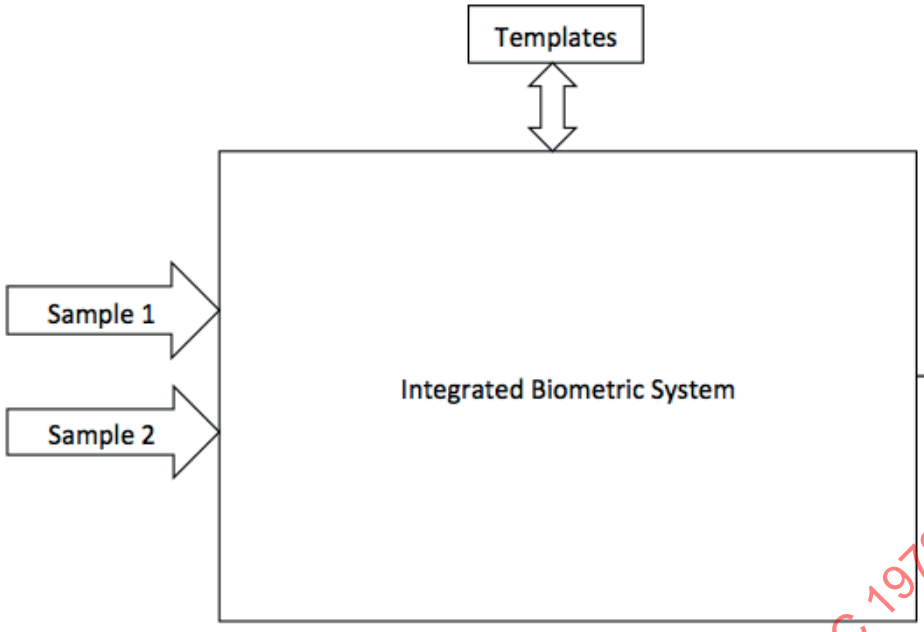

Figure F.2 — Fusion in an integrated multimodal system with decision output

F.2.1.2 Technology evaluation

Requirements for repeatability of decision-level fusion technology evaluation results are as follows:

— the decision fusion logic shall be identical;

— the function configurations (i.e. feature extraction, comparison, and decision) of Sub System 1 and Sub System 2, respectively, shall remain consistent across all tests;

NOTE 1 Sub System 1 and Sub System 2 can have different function configurations, and user-specific thresholds can differ for different users.

— the combination of Sample 1 and Sample 2 fed into each feature extraction function shall be identical;

— the combination of Template 1 and Template 2 shall be identical.

Consistent data selection methods for samples and templates are also required for evaluation repeatability.

If Sub System 1 and Sub System 2 are independent and separate, the evaluation report should include the following:

— identifying information for Sub System 1 and Sub System 2;

— identifying information for decision fusion logic;

— fusion level.

NOTE 2 See 6.4.2.

F.2.1.3 Scenario evaluation

Requirements for repeatability of decision-level fusion evaluation results are as follows:

— the function configurations (i.e. capture, feature extraction, comparison, and decision) of Sub System 1 and Sub System 2, respectively, shall remain consistent across all tests;

NOTE 1 Sub System 1 and Sub System 2 can have different function configurations, and user-specific thresholds may differ for different users.

— the decision fusion logic shall be identical;

— the combination of Sample 1 and Sample 2 fed into each feature extraction function shall be based on the same subject and position (e.g. right iris);

— the combination of Template 1 and Template 2 shall be based on the same subject and position.

Consistent data selection methods for samples and templates are also required for evaluation repeatability.

If Sub System 1 and Sub System 2 are independent and separate, the evaluation report should include the following:

— identifying information for Sub System 1 and Sub System 2;

— identifying information for decision fusion function;

— fusion level.

NOTE 2 See 7.4.2.

F.2.2 Score-level fusion

F.2.2.1 General

Fusion on the score level is illustrated in Figure F.3. Score-level fusion systems utilize score results from separate biometric subsystems.

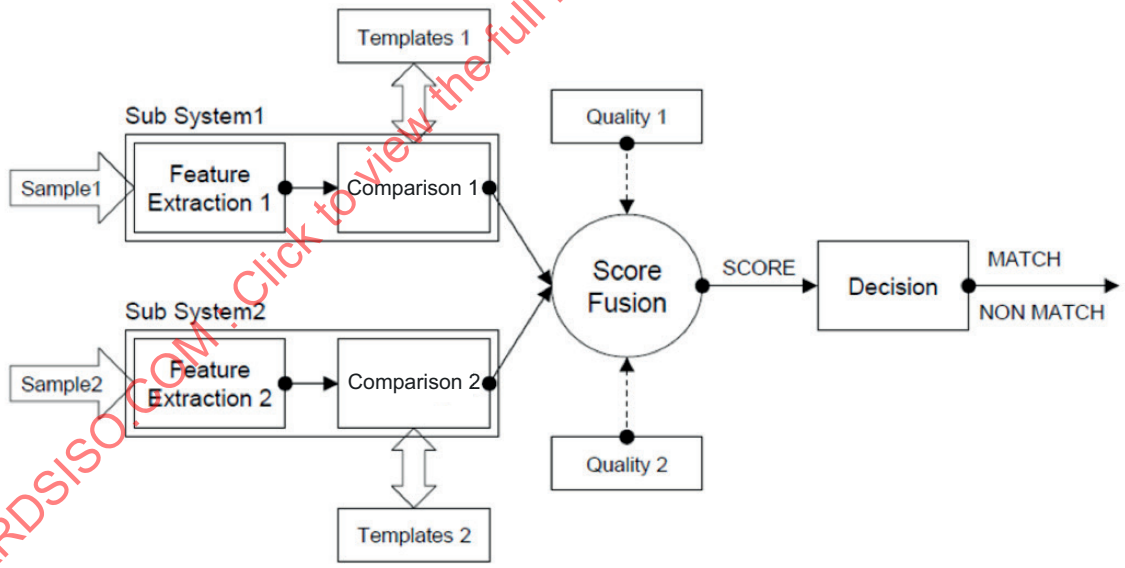

Figure F.3 — Score-level fusion

Score-level fusion may use sample quality in scenario or technology evaluations.

NOTE 1 Score-level fusion systems might be used to improve false match rate (FMR) and false non-match rate (FNMR).

Figure F.4 is an example of an integrated multimodal system with a score output in which insight is not provided into the implementation. This type of system can be tested without methods described in this annex.

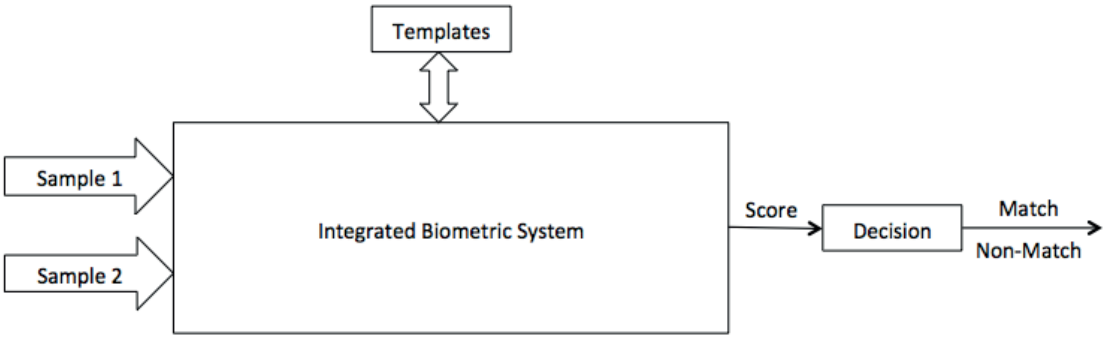

Figure F.4 — Fusion in an integrated multimodal system with score output

F.2.2.2 Technology evaluation

Requirements for repeatability of score-level fusion technology evaluation results are as follows:

— the score fusion function and decision function shall be identical;

— the function configurations (i.e. feature extraction and comparison) of Sub System 1 and Sub System 2, respectively, shall remain consistent across all tests;

NOTE 1 Sub System 1 and Sub System 2 can have different function configurations, and user-specific thresholds might differ for different users.

— the combination of Template 1 and Template 2 shall be identical;

— the combination of Sample 1 and Sample 2 fed into each feature extraction function shall be identical.

Requirements will be necessary for the data selection method for samples and templates, in order to keep repeatability.

If Sub System 1 and Sub System 2 are independent and separate, the evaluation report should include the following:

— identifying information for Sub System 1 and Sub System 2;

— identifying information for score fusion function and decision function;

— fusion level.

NOTE 2 See 6.4.2.

F.2.2.3 Scenario evaluation

Requirements for repeatability of score-level fusion scenario evaluation results can be stated as follows:

— the function configurations (i.e. capture, feature extraction, and comparison) of Sub System 1 and Sub System 2, respectively, shall remain consistent across all tests;

NOTE 1 Sub System 1 and Sub System 2 can have different function configurations, and user-specific thresholds can differ for different users.

— the score fusion function and decision function shall be identical;

— the combination of Template 1 and Template 2 shall be based on the same subject and position;

— the combination of Sample 1 and Sample 2 fed into each feature extraction function shall be based on the same subject and position.

Consistent data selection methods for samples and templates are also required for evaluation repeatability.

If Sub System 1 and Sub System 2 are independent and separate, the evaluation report should include the following:

— identifying information for Sub System 1 and Sub System 2;

— identifying information for score fusion function and decision function;

— fusion level.

NOTE 2 See 7.4.2.

F.2.3 Feature-level fusion

F.2.3.1 General

Fusion on the feature level is illustrated in Figure F.5. Feature-level fusion systems utilize results from separate feature extraction components.

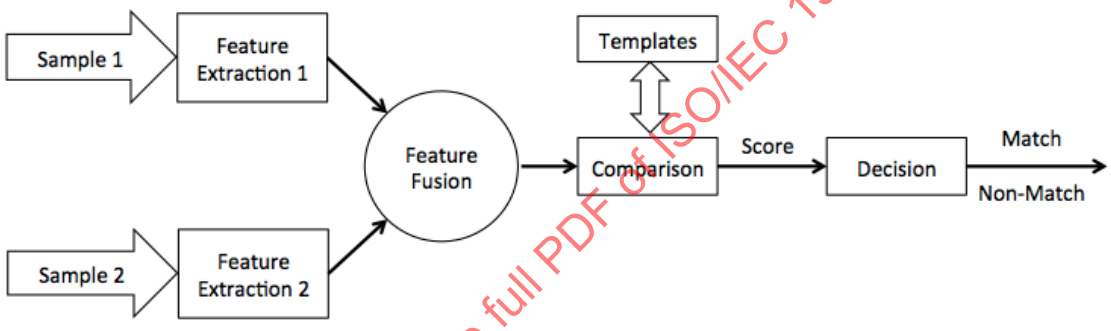

Figure F.5 — Fusion on the feature level

F.2.3.2 Technology evaluation

Requirements for repeatability of feature-level fusion technology evaluation results are as follows:

— feature fusion function, comparison function, and decision function shall be identical;

— the function configurations (i.e. feature extraction) of Feature extraction 1 and Feature extraction 2, respectively, shall remain consistent across all tests;

NOTE 1 Feature extraction 1 and Feature extraction 2 can have different function configurations, and user-specific thresholds can differ for different users.

— the combination of Sample 1 and Sample 2 fed into each feature extraction function shall be identical;

— the combination of Sample 1 and Sample 2 at the time of template creation shall be identical.

Consistent data selection methods for samples and templates are also required for evaluation repeatability.

If Feature extraction 1 and Feature extraction 2 are independent and separate, the evaluation report should include the following:

— identifying information for Feature extraction 1 and Feature extraction 2;

— identifying information for feature fusion function, comparison function and decision function;

— fusion level.

NOTE 2 See 6.4.2.

F.2.3.3 Scenario evaluation

Requirements for repeatability of feature-level fusion scenario evaluation results are as follows:

— feature fusion function, comparison function, and decision function shall be identical;

— the function configurations (i.e. capture and feature extraction) of Feature extraction 1 and Feature extraction 2, respectively, shall remain consistent across all tests;

NOTE 1 Feature extraction 1 and Feature extraction 2 can have different function configurations, and user-specific thresholds can differ for different users.

— the combination of Sample 1 and Sample 2 fed into each feature extraction function shall be based on the same subject and position;

— the fused sample template production process shall be identical.

Consistent data selection methods for samples and templates are also required for evaluation repeatability.

If Feature extraction 1 and Feature extraction 2 are independent and separate, the evaluation report should include the following:

— identifying information for Feature extraction 1 and Feature extraction 2;

— identifying information for feature fusion function, comparison function and decision function;

— fusion level.

NOTE 2 See 7.4.2.

F.2.4 Sample-level fusion

F.2.4.1 General

Fusion on the sample level is illustrated in Figure F.6. Sample-level fusion systems utilize samples from separate capture systems.

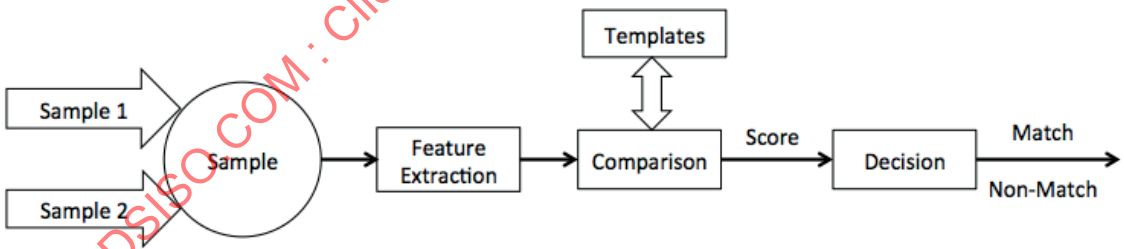

Figure F.6 — Fusion on the sample level

F.2.4.2 Technology evaluation

Requirements for repeatability of sample-level fusion technology evaluation results are as follows:

— sample fusion function, feature extraction function, comparison function and decision function shall be identical;

— the combination of Sample 1 and Sample 2 fed into each sample fusion function shall be identical;

— the combination of Sample 1 and Sample 2 at the time of template production shall be identical.

Consistent data selection methods for samples and templates are also required for evaluation repeatability.

NOTE See 6.4.2.

F.2.4.3 Scenario evaluation

Requirements for repeatability of sample-level fusion scenario evaluation results are as follows:

— the function configurations (i.e. capture) shall remain consistent across all tests;

— sample fusion function, feature extraction function, comparison function and decision function shall be identical;

— the fusion sample gathering process shall be identical.

Consistent data selection methods for samples and templates are also required for evaluation repeatability.

NOTE See 7.4.2.

F.3 Sensor type and presentation type information for repeatable evaluation

F.3.1 General

Due to the interdependencies between sensor type and presentation type, the two areas are addressed in this clause.

Sensor types are classified as follows under Note b, ISO/IEC TR 24722, Table 2:

— acquiring biometric data of multiple modalities with a single integrated sensor;

— acquiring biometric data of multiple modalities with different sensors.

Presentation types are classified as follows in ISO/IEC TR 24722, as shown in Figure F.7:

— acquisition and simultaneous presentation of biometric data of multiple modalities with single input trial;

— acquisition and sequential presentation of biometric data of multiple modalities with multiple input trials.

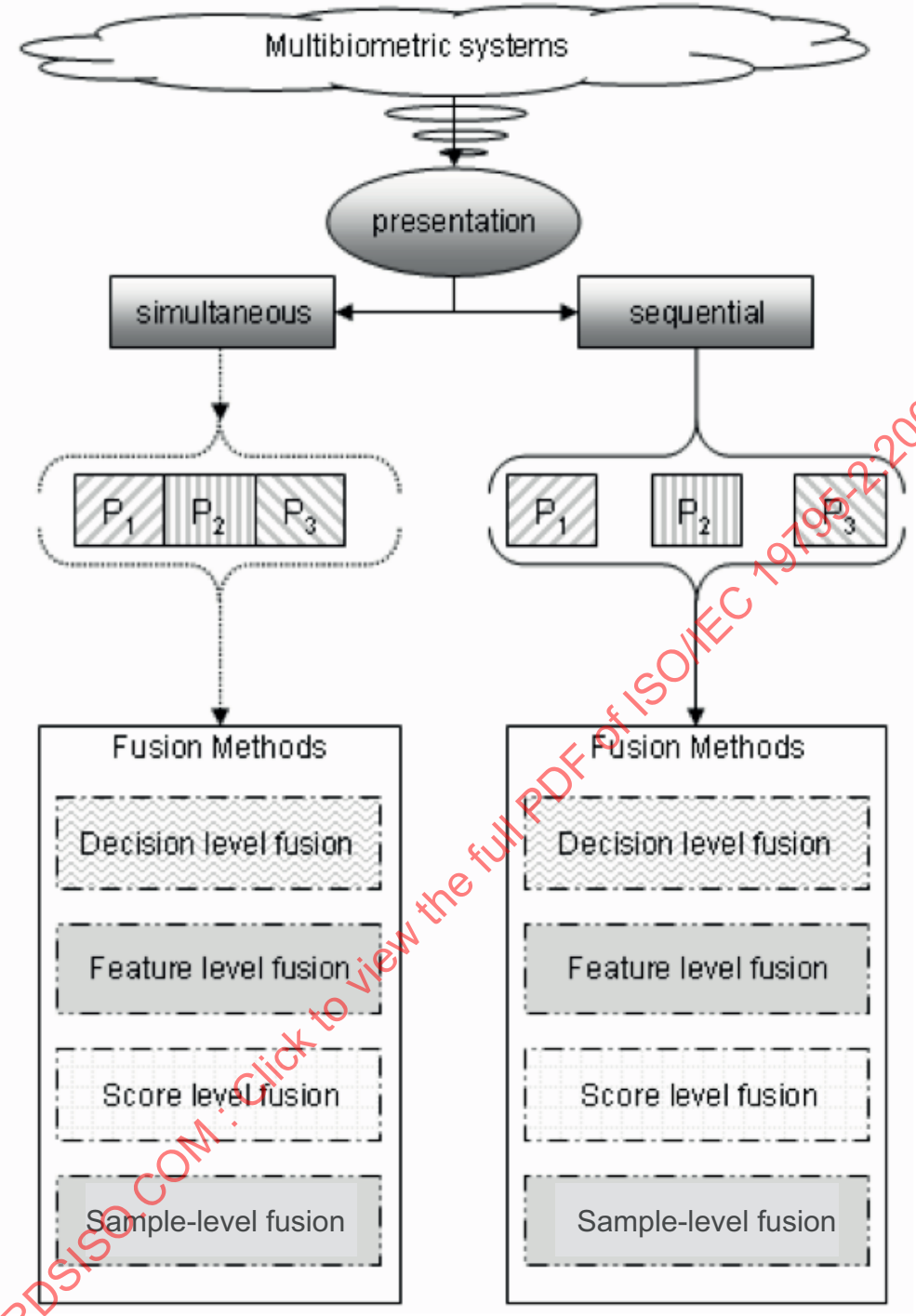

Figure F.7 — Types of presentation

F.3.2 Technology evaluation

Repeatability cannot be expected in evaluation results if the sensor type and presentation type differ during data collection. For this reason, the sensor and presentation type should be indicated clearly in the report.

NOTE See 6.4.2.

F.3.3 Scenario evaluation

Repeatability cannot be expected in evaluation results if the sensor type and presentation type differ. For this reason, the sensor and presentation type should be indicated clearly in the report.

NOTE See 7.4.2.

F.4 Decision-level fusion parameters for repeatable evaluation

F.4.1 General

ISO/IEC TR 24722 classifies decision-level fusion approaches as follows:

— simple decision-level fusion (e.g. AND fusion decision, or OR fusion decision);

— detailed decision-level fusion (i.e. layered or cascaded) as shown in Figure F.8.

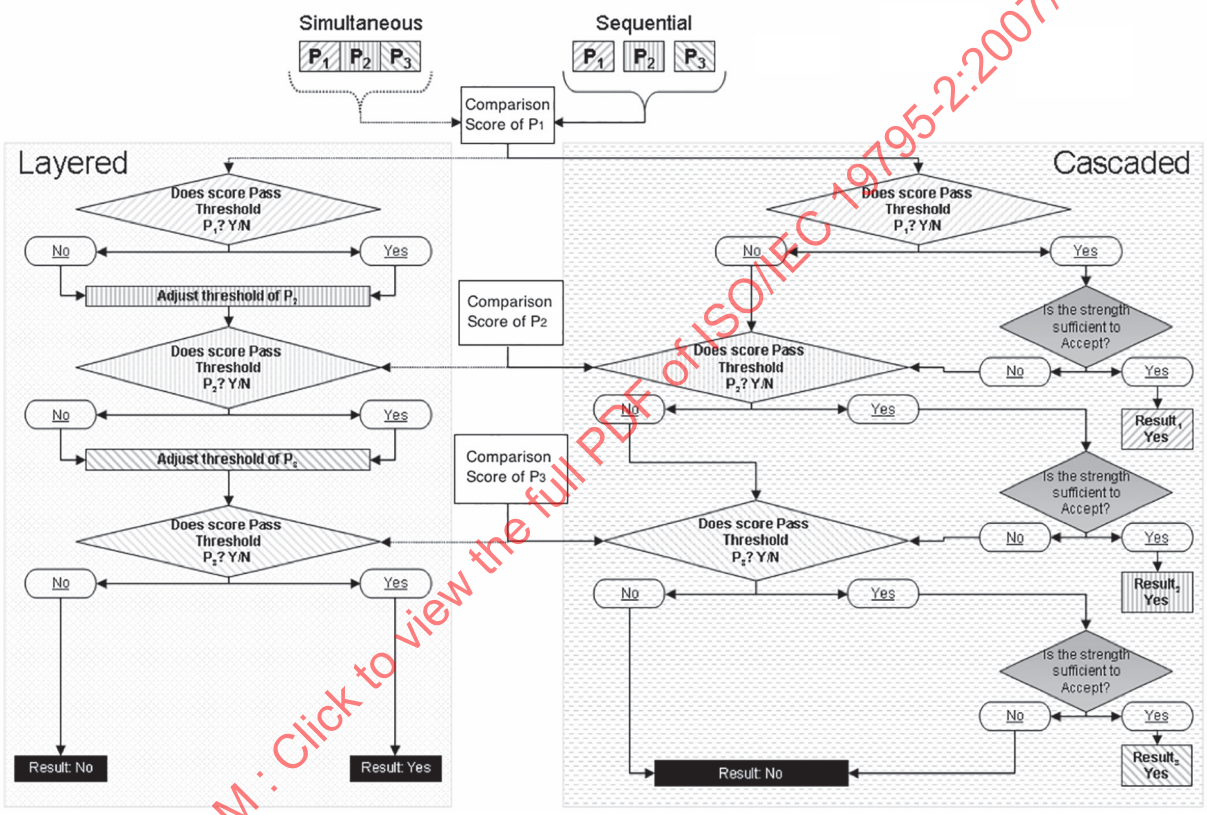

Figure F.8 — Detailed decision-level fusion

In the classification of these decision-level fusions, evaluation results vary by the parameters of the respective decision blocks. It is difficult to describe ROC/DET in evaluation results while changing multiple parameters simultaneously — for example, in a scenario evaluation of a multimodal system using sequential capture and cascaded decision level fusion logic. To execute a scenario test with decision-level fusion, each modality shall have an established threshold setting upon which the comparison decision is made. Since decision thresholds are in use, the opportunity to generate meaningful ROC/DET curves is limited, as results are single operating points (not the full ROC/DET across all thresholds).

When using Sequential presentation with Cascaded fusion logic, the comparison decision may be determined after one modality transaction (see Figure F.8, far right, first path to Result, Yes block.) If the Comparison decision is not determined after one modality transaction, then the second modality is presented (P2) and so on. Therefore, a distinctive element of multimodal evaluation reporting is to provide results at each step in the Cascade. To accomplish this, stage-based, tabular presentation of testing outcomes, such as those shown in Table F.1, should be used for this type of testing.

Table F.1 illustrates how evaluation results should be reported using N = 2 000 transactions (1 000 genuine and 1 000 imposter). Percentages in the table reflect the proportion of total, genuine or imposter transactions.

Table F.1 — Example of Scenario Evaluation Results Presentation for Decision-Level Fusion

Row		Number of modality transactions performed				Overall
		Stage 1	Stage 2		Stage 3	
1	Match decisions completed	N = 510	410	920 (cumulative)	90	1 010
2		25,5 %	20,5 %	46 %	4,5 %	50,5 %
3	Correct match (true accept rate)	500 (gen)	400	900	80	980
4		50 %	40 %	90 %	8 %	98 %
5	False match	10 (imp)	10	20	10	30
6		1 %	1 %	2 %	1 %	3 %
7	No match decision - false reject					1 000 - 980 = 20 2 %
8	Correct no match decision					1 000 - 30 = 970 97 %

Table F.1 shows the following.

— 500 of 1 000 genuine transactions correctly matched in Stage 1 (row 3, stage 1), while 10 imposters falsely matched in Stage 1 (row 5, stage 1).

— The columns for Stage 2 show number and percentage for Stage 2 alone (left column) as well as cumulative numbers and percentages for Stages 1 and 2 (right column). 410 of the 1 000 genuine transactions correctly matched in Stage 2 (row 3, stage 2), while 10 imposters falsely matched in Stage 2 (row 5, stage 2).

— The far right column shows the overall results for all 3 stages cumulatively. The conventional genuine comparison metrics (shaded green) show 98 % TAR and 2 % FRR. Conventional imposter overall metrics (shaded pink) show 3 % FMR and 97 % correct reject rate.

While transaction times and throughput are not quantitatively addressed by this table, as the number of first stage matches increases (row 1, stage 1), the throughput potential for the overall system improves.

This tabular reporting approach provides more insight into multimodal performace aspects than traditional summary statistics presented in the last column.

F.4.2 Technology evaluation and scenario evaluation

The evaluation report should identify the decision logic and all relevant parameters in each decision block when generating the ROC/DET.

NOTE See 6.4.

F.5 Fusion Information Format (FIF) value for repeatable evaluation

F.5.1 General

ISO/IEC 29159-1 defines the fusion information format (FIF), the format for initialization parameters for each modality, for decision-level fusion. Because the use of fusion information format (FIF) causes change in fusion model features, it will affect repeatability in evaluation.

F.5.2 Technology evaluation and scenario evaluation

