INTERNATIONAL STANDARD

ISO/IEC 29794-1

First edition
2009-08-01

Information technology — Biometric sample quality —

Part 1: **Framework**

Technologies de l'information — Qualité d'échantillon biométrique —

Partie 1: Cadre

Reference number
ISO/IEC 29794-1:2009(E)

PDF disclaimer

This PDF file may contain embedded typefaces. In accordance with Adobe's licensing policy, this file may be printed or viewed but shall not be edited unless the typefaces which are embedded are licensed to and installed on the computer performing the editing. In downloading this file, parties accept therein the responsibility of not infringing Adobe's licensing policy. The ISO Central Secretariat accepts no liability in this area.

Adobe is a trademark of Adobe Systems Incorporated.

Details of the software products used to create this PDF file can be found in the General Info relative to the file; the PDF-creation parameters were optimized for printing. Every care has been taken to ensure that the file is suitable for use by ISO member bodies. In the unlikely event that a problem relating to it is found, please inform the Central Secretariat at the address given below.

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents

Page

Foreword iv

Introduction v

1 **Scope** 1

2 **Conformance** 1

3 **Normative references** 1

4 **Terms and definitions** 2

5 **Acronyms and abbreviated terms** 5

6 **General biometric system** 5

7 **Biometric sample quality criteria** 5
7.1 **Reference model** 5
7.2 **Quality components: character, fidelity, utility** 6
7.3 **Usefulness of quality data** 7
7.3.1 **Real-time quality assessment** 7
7.3.2 **Use in different applications** 7
7.3.3 **Use as a survey statistic** 7
7.3.4 **Accumulation of relevant statistics** 8
7.3.5 **Reference dataset improvement** 8
7.3.6 **Quality-based conditional processing** 8
7.3.7 **Interchange of quality data by disparate systems** 8

8 **Data interchange format field definition** 9
8.1 **Data fields** 9
8.2 **Quality score** 9
8.2.1 **Purpose** 9
8.2.2 **Data transformation considerations** 10
8.2.3 **Failure modes** 10
8.2.4 **Resolution** 10
8.2.5 **Summarization** 10
8.3 **Quality algorithm identification (QAID)** 10
8.3.1 **Overview** 10
8.3.2 **Methodology** 10
8.4 **Standardized exchange of quality algorithm results** 11

9 **Normalization** 12

Annex A (informative) **Procedures to construct a quality score normalization dataset** 13

Annex B (informative) **Example - standardized exchange of quality algorithm results** 19

Annex C (informative) **Procedures for aggregation of utility-based quality scores** 21

Bibliography 23

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 29794-1 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 37, *Biometrics*.

ISO/IEC 29794 consists of the following parts, under the general title *Information technology — Biometric sample quality*:

— *Part 1: Framework*

— *Part 4: Finger image data* [Technical Report]

— *Part 5: Face image data* [Technical Report]

Future parts of ISO/IEC 29794 will address other modalities specified by ISO/IEC 19794, with part numbers and titles aligned appropriately. However, as ISO/IEC 29794-1 is intended for use by all modalities, a modality does not necessarily need a modality-specific part in order to make use of quality scores.

It is anticipated that a future version of each part of ISO/IEC 19794 will normatively reference ISO/IEC 29794-1, and their respective data fields will be updated as required.

Introduction

Quality metrics are useful for several applications in the field of biometrics. ISO/IEC 19784-1 specifies a structure and gives guidelines for quality score categorization, and ISO/IEC 29794 defines and specifies methodologies for objective, quantitative quality score expression, interpretation, and interchange. This part of ISO/IEC 29794 is intended to add value to a broad spectrum of applications in a manner that

a) encourages competition, innovation, interoperability and performance improvements; and

b) avoids bias towards particular applications, modalities, or techniques.

This part of ISO/IEC 29794 presents several biometric sample quality scoring tools, the use of which is generally optional but can be determined to be mandatory by particular application profiles or specific implementations.

A number of applications can benefit from the use of biometric sample quality data; an example is the use of real-time quality feedback upon enrolment to improve the operational efficiency and performance of a biometric system. The association of quality data with biometric samples is an important component of quality metric standardization. Quality fields as specified in 8.1 will be incorporated into data interchange formats. If a CBEFF header is present, then CBEFF_BDB_quality can additionally be used to express quality data. Useful analyses can be performed using quality data along with other data in order to improve the performance of a biometric system. For example, correlating quality data to other system metrics can be used to diagnose problems and highlight potential areas of performance improvement.

Information technology — Biometric sample quality —

Part 1:
Framework

1 Scope

For any or all biometric sample types as necessary, this part of ISO/IEC 29794

1. establishes terms and definitions that are useful in the specification, and use of quality metrics;
2. recommends the purpose and interpretation of biometric quality scores;
3. defines the format and placement of quality data fields in biometric data interchange formats;
4. suggests methods for developing biometric sample datasets for the purpose of quality score normalization; and
5. suggests a format for exchange of quality algorithm results.

Outside the scope are the following:

1. the specification of minimum requirements for sample, module, or system quality scores;
2. performance assessment of quality algorithms; and
3. standardization of quality algorithms.

2 Conformance

A block of quality data is in conformity with this part of ISO/IEC 29794 if it conforms to the normative requirements of Clause 8.

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.

19794-1:2006, *Information technology — Biometric data interchange formats — Part 1: Framework*

19785-2:2006, *Information technology — Common Biometric Exchange Formats Framework — Part 2: Procedures for the operation of the Biometric Registration Authority*

4 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

4.1
acquisition fidelity
fidelity of a sample attributed to the acquisition process

4.2
biometric failure to enrol
failure of the biometric system to store a usable biometric reference due to deficiencies in the biometric data during an enrolment application

NOTE 1 Deficiencies in the biometric data can result from failure to capture an adequate or usable biometric sample, failure to extract adequate or usable biometric features from the sample, or failure to generate an adequate or usable biometric reference from the biometric features.

NOTE 2 See SC 37 N SD2 for most recent definition.

4.3
biometric failure to enrol rate
proportion of biometric enrolment sessions that resulted in a biometric failure to enrol for other than non-biometric reasons

NOTE 1 Basing the denominator on the number of biometric enrolment sessions can result in a higher value than basing it on the number of biometric capture subjects.

NOTE 2 The proportion denominator is the number of biometric enrolment sessions, excluding those sessions that failed to complete for non-biometric reasons.

NOTE 3 See SC 37 N SD2 for most recent definition.

4.4
character
contributor to quality of a sample attributable to inherent features of the source

4.5
environment
physical surroundings and conditions where biometric capture occurs, including operational factors such as operator skill and enrolee cooperation level

4.6
extraction fidelity
component of the fidelity of a sample attributed to the biometric feature extraction process

4.7
extrinsic
〈quality score〉 requiring reference to an external source, such as a standard, register, or technical specifications, for full interpretation and normalization

4.8
fidelity
expression of how accurately a biometric sample represents its source biometric characteristic

NOTE The fidelity of a sample comprises components attributable to one or more of the processing steps: acquisition, extraction, signal processing.

4.9
intrinsic
⟨quality score⟩ conveying fully interpreted, normalized data without the requirement for additional extrinsic information for quality score normalization

4.10
interpretation
process of analyzing a quality score along with other data in order to give that score contextual, relative meaning

4.11
failure to acquire rate
proportion of the biometric application attempts that resulted in failure to acquire an adequate or usable biometric sample, for other than non-biometric reasons

NOTE 1 The proportion denominator is the number of biometric enrolment attempts, excluding those attempts that failed to complete for non-biometric reasons.

NOTE 2 See SC 37 N SD2 for most recent definition.

4.12
false match rate
FMR
proportion of the completed biometric non-match trials that result in a false match

NOTE 1 The value computed for the false match rate will depend on thresholds, and other parameters of the comparison process, and the protocol defining the biometric non-match trials. In particular, treatment of comparisons between

— identical twins,

— completely different biometric characteristics of different individuals, such as face topography and Galton ridges, and

— different but related biometric characteristics from the same individual, such as left and right hand topography,

will need proper consideration. See ISO 19795-1.

NOTE 2 “Completed” refers to the computational processes required to make a comparison decision, i.e. failures to decide are excluded.

NOTE 3 See SC 37 N SD 2 for most recent definition.

4.13
false non-match rate
FNMR
proportion of the completed biometric match trials that result in a false non-match

NOTE 1 The value computed for the false non-match rate will depend on thresholds and other parameters of the comparison process, and the protocol defining the biometric match trials.

NOTE 2 “Completed” refers to the computational processes required to make a comparison decision, i.e. failures to decide are excluded.

NOTE 3 See SC 37 N SD2 for most recent definition.

4.14
operator
individual who processes a user in a biometric system, performing or supervising capture and recapture

4.15
performance
assessment of the FMR, FNMR, failure to enrol rate and failure to acquire rate of a biometric system

4.16
quality
degree to which a biometric sample fulfils specified requirements for a targeted application

NOTE Specified quality requirements can address aspects of quality such as focus, resolution, etc. Implicit quality requirements address the likelihood of achieving a correct matching result.

4.17
quality score
quantitative expression of quality

4.18
quality score normalization
rescaling of quality scores to improve consistency in scale and interpretation

4.19
quality score normalization dataset
QSND
dataset of biometric samples annotated with quality scores for use in quality score normalization

NOTE Target quality scores can be assigned on the basis of performance outcomes using the sample in question, or can be based on quality factors recorded in acquisition of the dataset.

4.20
quality score percentile rank
QSPR
percentile rank of the quality score of a biometric sample, derived from its own utility score and those of other samples in an identified control dataset

cf. quality score normalization dataset

4.21
raw quality score
quality score that has not been interpreted, either by the creator or recipient of the score, and alone can not intrinsically provide contextual information

4.22
sample
image, signal, or pattern based interpretation of a physical human feature used for identification or verification using biometric techniques

4.23
source
physical body part or function represented by a biometric sample

4.24
utility
observed performance of a biometric sample or set of samples in one or more biometric systems

NOTE 1 The character of the sample source and the fidelity of the processed samples contribute to – or similarly detract from – the utility of the sample.

NOTE 2 Utility can combine performance measures such as FMR, FNMR, failure to enrol rate, and failure to acquire rate.

5 Acronyms and abbreviated terms

BDB biometric data block

BIR biometric information record

CBEFF common biometric exchange formats framework (ISO/IEC 19785)

FERET facial recognition technology database

FMR false match rate

FNMR false non-match rate

QAID quality algorithm identification

QSND quality score normalization dataset

QSPR quality score percentile rank

XML extensible markup language

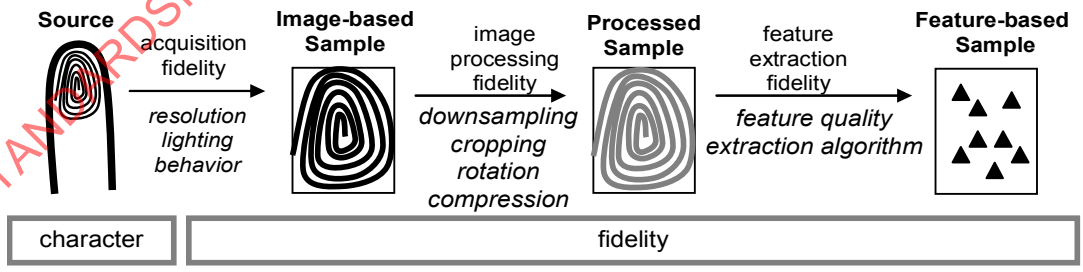
6 General biometric system

A general biometric system is described in Standing Document 11, ISO/IEC JTC 1/SC 37 Part 1 Overview Standards Harmonization Document (SC 37 N-SD11).

7 Biometric sample quality criteria

7.1 Reference model

In biometrics, the term “quality” is used to describe several different aspects of a biometric sample that contribute to the overall performance of a biometric system. For the purposes of standardization, this document defines terms, definitions, and a reference model for distinguishing between these different aspects of quality, illustrated in Figure 1. Figure 2 illustrates the relationship between character, fidelity, quality, utility, and system performance.



Quality = Function [character, fidelity components]
Utility reflects the impact of the quality of a single sample on system performance

Figure 1 — Quality reference model illustration

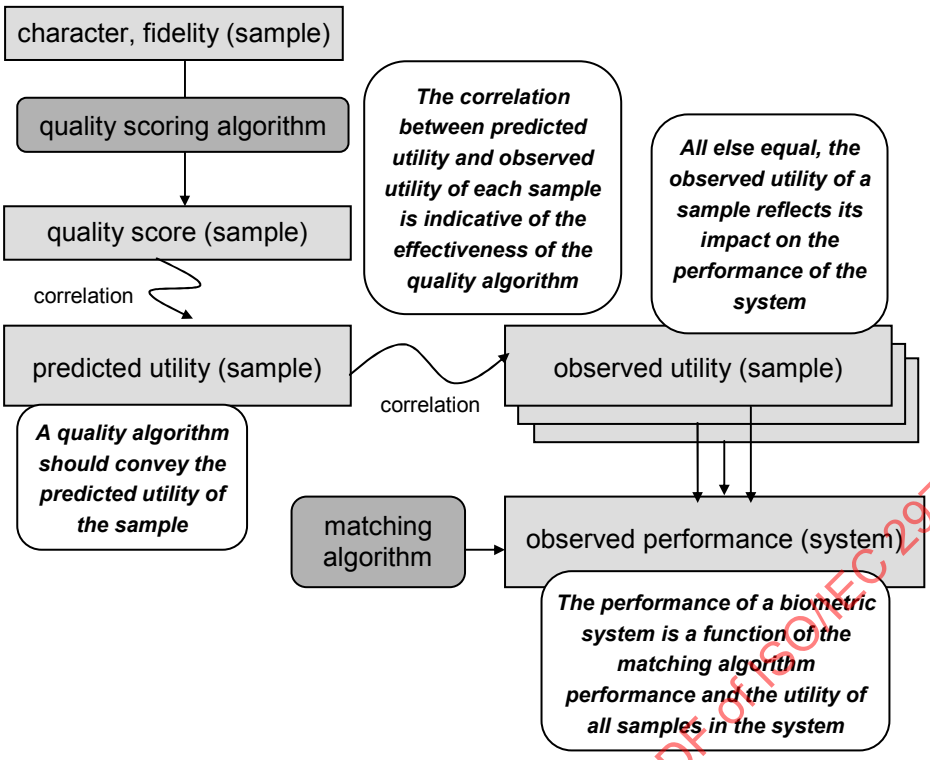


Figure 2 — Relationship between quality and system performance

7.2 Quality components: character, fidelity, utility

The term "quality" as it is currently used in the field of biometrics has several connotations, depending on context. Three prevalent uses are to subjectively reflect:

1. the ***character*** of a sample. An expression of quality based on the inherent features of the source from which the biometric sample is derived. For example, a scarred fingerprint has poor character, and blepharoptosis (droopy eyelid) causes poor iris character;

2. the ***fidelity*** of a sample to the source from which it is derived. An expression of quality based on fidelity reflects the degree of its similarity to its source. Sample fidelity is comprised of fidelity components contributed by different processes;

3. the ***utility*** of a sample within a biometric system. An expression of quality based on utility reflects the predicted positive or negative contribution of an individual sample to the overall performance of a biometric system. Utility-based quality is dependent on both the character and fidelity of a sample. Utility-based quality is intended to be more predictive of system performance, e.g. in terms of FMR, FNMR, failure to enrol rate, and failure to acquire rate, than measures of quality based on character or fidelity alone. (See Table 1)

The term "quality" should not be solely attributable to the acquisition settings of the sample, such as image resolution, dimensions in pixels, grayscale/color bit depth, or number of features. Though such factors may affect sample utility and could contribute to the overall quality score.

Note that the character and utility of an acquired sample depend on the features to be considered by the comparator. For instance, the same finger image may be of low character and utility with respect to minutiae recognition (because of too few minutiae), but of high character and utility with respect to spectral pattern recognition.

Table 1 — Illustration of relationship between fidelity, utility, and character

		Fidelity	
		Low	High
Character	Low	Low fidelity and low character results in low utility. Recapture might improve utility. However, if possible use of other biometric characteristics is recommended.	High fidelity and low character results in low utility. Recapture will not improve utility. Use of other biometric characteristics is recommended.
	High	Samples with high character and low fidelity typically will not demonstrate high utility. Utility can be improved upon recapture or image enhancement techniques.	Samples with high character and high fidelity indicate capture of useful sample. High utility is expected.

7.3 Usefulness of quality data

7.3.1 Real-time quality assessment

Real-time quality data can be used by an operator, automated system, or user to help improve the average quality of biometric samples submitted upon enrolment. This feedback might indicate the character, fidelity, utility, and improvability of a sample. In this way, operational efficiency and overall system performance can be improved by assisting an operator, or augmenting an automated quality control system, in decisions to a) accept the sample, b) reject the sample, c) reattempt a capture, or d) declare a failure to acquire or failure to enroll. Quality data can be retained for later use in, for example, determining whether an enrolment sample should be replaced when the next sample is captured.

7.3.2 Use in different applications

A particular biometric sample might be used for several different applications and therefore its associated quality data should be applicable to all of these. This would include both one-to-one and one-to-many comparisons involving the use of comparison algorithms from different vendors that would interpret sample features differently and yield different comparison scores. The challenge in establishing a universal quality standard is in defining a metric that is sufficiently adaptable for use by all applications for which a particular sample might be used given that metrics of utility vary greatly between applications. Therefore, it should be recognized that it is a technical challenge to define a singular metric that accurately conveys the utility of a biometric sample for all applications for which it may be used, and this should be taken into consideration in defining quality standards. Thus a quality metric—ideally predicting performance for a comparator or class of comparators —will likely be designed to capture only some of the failure modes and sensitivities of only a limited number of biometric systems. It may be useful to apply more than one quality metric in order to improve predictability of various failure modes.

It is useful for recipients of quality score data to be able to differentiate between scores generated by different quality algorithms and capture equipment. This data may be used to enable recipient software to be configured so that different thresholds or quality classifications can be applied to scores generated by different algorithms. In addition, by differentiating between scores from different algorithms, a recipient may isolate results from different algorithms and use the data to optimize thresholds accordingly.

7.3.3 Use as a survey statistic

Quality scores may be used to monitor operational quality. For example aggregated quality scores could be compared with preset limits or monitored against an operational requirement. If, for example, quality scores are generated from biometric samples collected at many sites, or over different time periods, then they may be used to identify anomalous operation. For example, if face image quality is computed at the license issuance

desks at a Department of Motor Vehicles, then a ranked list of aggregated quality scores could be used to identify desks that exhibit a lower than average quality, or to monitor trends over weeks or months.

7.3.4 Accumulation of relevant statistics

Reliable quality scores may be used to survey users and transactions to accumulate statistics giving conditional probabilities of the kind “given a quality X sample on finger A, what is the likelihood of a quality Y sample from finger A (or finger B)". This will inform the system and/or operators over whether a higher quality sample is likely if another capture is attempted.

7.3.5 Reference dataset improvement

The association of quality data with a sample that is to be entered into a reference dataset is important for the maintenance and improvement of the reference dataset quality. The tracking of sample quality can lead to detection of potential deterioration of operator training or it may indicate deterioration in the performance of the sample capture equipment. Tracking of the sample quality data should be an important part of the biometric system’s operating procedures. The quality data may also be used to improve the quality of the reference file, and hence the performance of the biometric system. Improvement can be made by the replacement or possible augmentation of the stored information so as to make use of the best available quality data. Typically, the replacement decisions are linked to the comparator performance of the system processing the data.

7.3.6 Quality-based conditional processing

Biometric samples can be processed differently based on quality metrics. In particular, poor-quality data can be processed using different algorithms or thresholds than normal.

7.3.7 Interchange of quality data by disparate systems

Standardized exchange of quality data between disparate systems is useful in retaining the modular interchangeability of local or remote system hardware and software components, and the integrity of quality data in the event of such an interchange.

For example, by using standardized exchange of quality data, consumers of quality data from a component require minimal modification if that component is replaced.

8 Data interchange format field definition

8.1 Data fields

Table 2 summarizes the structure of a biometric data quality block.

Table 2 — Data fields

description		size	valid values	notes
Number of Quality Blocks		1 byte	[0,255]	This field is followed by the number of 5-byte Quality Blocks reflected by its value (see Figure 3). A value of zero (0) means that no attempt was made to assign a quality score. In this case, no Quality Blocks are present.
Quality Block	Quality Score	1 byte	[0,100] 255	0: lowest 100: highest 255: failed attempt to assign a quality score
	Quality Algorithm Vendor ID	2 bytes	[1,65535]	Quality Algorithm Vendor ID shall be registered with IBIA as a CBEFF biometric organization. Refer to CBEFF vendor ID registry procedures in ISO/IEC 19785-2.
	Quality Algorithm ID	2 bytes	[1,65535]	Quality Algorithm ID may be optionally registered with IBIA as a CBEFF Product Code. Refer to CBEFF product registry procedures in ISO/IEC 19785-2.

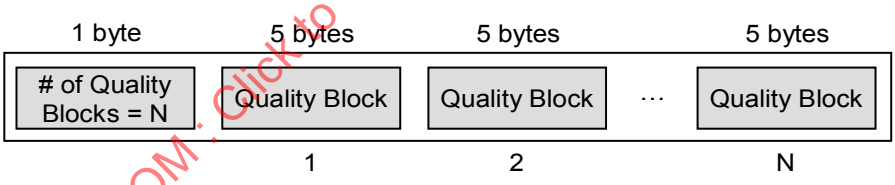


Figure 3 — Structure of quality field

Quality scores should always be placed within the quality score field of the biometric data block (BDB) as defined in ISO/IEC19794-x associated with the sample. CBEFF quality fields should not be used in place of 19794 quality fields but rather as supplementary data. The prescribed use of CBEFF quality fields may be supplied by each CBEFF patron format standard and is beyond the scope of this document. Note that multiple quality scores calculated by the same algorithm (same algorithm ID) shall not be present in a single BDB.

8.2 Quality score

8.2.1 Purpose

Quality algorithms shall produce quality scores that predict performance metrics such as either false match or false non-match. In cases where the system utilizes components from multiple vendors, the quality scoring method should aim to reflect the aspects of performance important for each algorithm used. As noted in 7.3.2, it is challenging to find a single quality measure that is universal, not vendor-specific and yet adequately indicates performance, and it may be useful to apply more than one quality algorithm.

8.2.2 Data transformation considerations

Data transformation by an application system is likely to impact the data quality (ie. down-sampling or further compression). The impact of such transformations on the data quality metrics may be recomputed by the application system in accordance with guidance provided by this standard. Any time a biometric sample undergoes a transformation, the quality of the transformed sample should be reassessed and associated with the transformed sample. For example, throughout an identity management system a biometric sample may be stored in multiple formats (e.g., high resolution fingerprint image stored centrally and a minutia-based representation stored on a smart card).

8.2.3 Failure modes

To be predictive of performance it may benefit a quality algorithm designer to produce quality scores that are intended to model known failure modes / sensitivities of a biometric comparator and image/signal processing algorithms. Further, to achieve some measure of generality the quality score should be based on the set of sensitivities that are common to a class of system (e.g. minutia comparators).

8.2.4 Resolution

A quality apparatus shall provide for a mapping to at least four discrete values, which, when utilized towards a variety of applications, still maintains the ability to discriminate between distinct levels of performance, such as "excellent", "adequate", "marginal", and "unacceptable".

8.2.5 Summarization

Annex C suggests procedures for the appropriate aggregation of utility-based quality scores over a collection of samples, e.g. enterprise-wide summarization. The result is a summary value which supports monitoring of quality. Quality summarization should be performed across similar usage, e.g. quality summarization over all enrolment samples of an enterprise, or quality summarization over all verification samples of an enterprise. In operations where users frequently interact with a biometric system (e.g. time and attendance applications), quality scores may be aggregated on a per user basis. This will reveal the existence of individuals that consistently yield low quality samples.

8.3 Quality algorithm identification (QAID)

8.3.1 Overview

The Quality Algorithm ID (QAID) is an identifier of the quality algorithm used to calculate the quality score of the sample. As long as there are no common criteria for quality assessment, it is indispensable to enable the recipient of a BIR to differentiate between quality scores generated by different quality algorithms and adjust for any differences in processing or analysis as necessary. A QAID Registry would provide a reference that indicates the vendor and version number of the identified quality algorithm. The QAID method is considered to be a solution that may be implemented quickly but only partially achieves the goals of quality score standardization.

8.3.2 Methodology

This method requires as normative that:

If no quality scoring is attempted, then the value of the Number of Quality Blocks field is 0 and there are no Quality Blocks present. If Number of Quality Blocks is between 1 and 255, then an identifier of the quality algorithm used to generate the score shall be present using the Quality Algorithm Vendor ID and Quality Algorithm ID fields according to Table 2 — Data fields. Note that this method does not preclude, but rather complements, further work to standardize a universal quality scoring method (i.e. a score that intrinsically includes some degree of normalization) such as QSND. See Clause 9.

A feature of this “quality algorithm identification” method is that the recipient of the raw quality score data may need to do some local analysis and/or processing to fully interpret the meaning of the scores. In other words, the sender of the score is not attempting to interpret the quality score for a potentially unknown application or destination. But importantly, the recipient can obtain the information on how the quality score is established

from the quality algorithm vendor and develop appropriate means to automatically distinguish between quality scores generated by different quality algorithms, and interpret them appropriately.

8.4 Standardized exchange of quality algorithm results

Quality algorithm vendors should be able to offer results of their quality algorithms in a standardized way to the biometric community. On the other hand, consumers of ISO/IEC 19794 data interchange records are able to retrieve and process this information effectively in order to assess the value of the output of this quality algorithm to their implementation. This approach has the following benefits:

- Both, quality algorithm vendors as well as consumers have the ability to gain value from technical improvements, which is a necessary prerequisite in the starting phase of wide spread quality score use.
- In some applications, updates may be retrieved automatically, if the necessary infrastructure is there.
- It will re-shift the evaluation effort related with QAID from the consumer and integrators back to the quality algorithm vendors (which do the evaluation anyway).
- Over time, standardized test sets will evolve,
  - as it is in the interest of the quality algorithm vendor to use (a) reporting test set(s), that is of use for many costumers, and
  - the need for new test sets will vanish over time and the use of such test sets will be critically reviewed by the community.
- Evolution of test sets will facilitate the development of QSND.

For the exchange, the following items shall be provided:

1. the quality algorithm vendor ID,
2. the quality algorithm ID,
3. the minimum and maximum theoretical output value of the algorithm,
4. the unique name of test set used (e.g. in form of "FERET-Grayscale" in the case of face recognition), and
5. the list of samples (e.g. for FERET “Duplicate 1” in the case of face recognition) that have been processed.

Everyone will be able to publish new test sets (biometric samples + a naming scheme).

A self describing language like XML will be used for the description of the data sets as well as for the evaluation results. The evaluation results could be maintained in a central registry or on a vendor site (via a link in the central registry).

An example implementation using XML can be found in Annex B.

9 Normalization

Normalization of quality score data is the process by which quality score data is processed by its recipient in order to give the scores local context and meaning, such as to make quality scores from different algorithms have similar meaning.

A raw quality score is assigned to a biometric sample by a particular quality algorithm. In order to interpret the raw score, the recipient of a score must have some contextual information. This information may be provided:

1. extrinsically in the form of metadata or off-line data (e.g. standard) that instructs the recipient on interpretation of the score. For example, if a quality score is accompanied by identification of the algorithm used to generate the quality score of the associated sample (i.e. QAID), then recipient software can be configured to use vendor-supplied data (e.g. suggested thresholds) to best process the sample. The algorithm could alternatively be used to perform analysis in order to fully optimize the interpretation of the scores given the local application and data. By identifying the algorithm, scores created by different algorithms could be differentiated so that, for example, different thresholds could be applied to the sample depending on the source of the quality score.

2. intrinsically, in the form of a normalized quality score. Normalization of quality score data provides contextual information about the score. An example is a quality score representing the perceived likelihood that a sample, when matched, will result in a false non-match.

QAID enables vendor-specific scaling, such that the 0-100 scale correlates to some other scale reflecting the above. For example, the recipient of a file would be encouraged to analyze the correlation of quality scores to FMR and FNMR of the samples processed by their comparator. The results could be used to, for example, specify an acceptance operating threshold. This method provides the recipient the information necessary to interpret the scores in a way that is relevant to their own environment and application, and permits the use of many different algorithms or versions of algorithms in a single system.

The purpose of quality score normalization (QSN) corpus is to provide a consistent interpretation of quality scores through normalizing quality scores or Quality Score Percentile Rank (QSPR). QSPR enables universal expression and interpretation of a quantitative sample quality score, which is that quality algorithm “X” considers biometric sample “Y” to have a quality percentile rank “Z”. The translation of raw quality scores to percentile rank scores is achieved by running a standardized corpus of samples through a given quality algorithm and pairing all possible raw score outcomes to percentile rank scores.

See Annex A for more information.

Annex A
(informative)

Procedures to construct a quality score normalization dataset

A.1 Quality score normalization dataset (QSND) overview

The Quality Score Normalization Dataset (QSND) standardization method aims to enable a consistent and interoperable interpretation of the quality score. To achieve this, the QSND method requires the standardization of a dataset or datasets, with each biometric modality having its own dataset(s). For each dataset, several score classes will be established, with each class being assigned with a range of quality scores. Several sample data that fall into each class will be collected and made available in the standards for the algorithm developers and recipients (users) of quality scores. The algorithm providers are then advised to modify or provide a transformation function such that the output of their quality algorithm will be consistent to the recommended range of score for each sample data provided in the standard database. However, if there would likely be more than one version of the standardized dataset, then it would be useful information to identify the dataset used for the quality score normalization. The dataset will then be made as part of the standard and will be made available for the algorithm developers and recipients (users) of quality scores.

This annex defines a procedure for constructing a performance-oriented reference database. The result is a set of samples annotated with a target quality score. The value is essentially a consensus similarity score classification from a set of comparators. Such reference sets are of primary use for normalization of quality score from different quality algorithms so that quality scores of various quality algorithms have a consistent interpretation within a reasonable and practical range. The same method would be of use to tune a quality algorithm to an operational situation in which the comparator and kind of data are known and available or by quality algorithm developers working on the general problem. The input to our procedure is a representative sample database. The output is an annotation of each sample with a scalar quality target. The method presumes the availability of a representative comparison algorithm, which will be used to compare samples to produce both genuine and impostor similarity scores. It is therefore implied that two or more samples per person are available.

A.2 Data

Data gathered in a target operational application would be most realistic. It is recommended to create a reference set with a larger proportion of samples that are naturally problematic to the comparator than is present in the population. Insofar as possible, it should be balanced; for example--in the case of fingerprints--in terms of finger position (right/left index/thumb/middle), finger impression (roll/plain/flat), sex, age, and capture device. Lack of data often renders it difficult to create such a balanced dataset.

In some areas, this method may ultimately rely on the standardization of biometric sample datasets, and their subsequent maintenance and availability. At a minimum, datasets should be maintained in a manner that provides 1) adequate space for file storage, 2) the protection of the integrity of the stored dataset files, and 3) secure, controlled access to files by qualified individuals.

A.3 Target quality assignment

We seek to assign a performance-based quality score to each image in a reference dataset. We ensure that the quality values are representative of performance by associating the image with similarity scores as follows. Consider a biometric dataset containing $N_i \geq 2$ samples, $d_i^{(1)}$, $d_i^{(2)}$,..., $d_i^{(N_i)}$, for each of M subject, $i = 1,...,M$ where each image contains only one biometric characteristic, i.e. image of just one finger or one

iris. The following procedure assigns quality values $q_i^{(1)}$, $q_i^{(2)}$,..., $q_i^{(N_i)}$ $i=1,...,M$ to all images in the reference dataset.

I. For each comparator V_k, $k=1,\ldots,K$, *of* K available comparators

For each instance record $d_i^{(u)}$ (i.e. the u^{th} sample of subject i):

1) Generate the set of all possible mated comparison scores using the k-th comparator,

$$\begin{aligned} S_{ii} &= \{s_{i,i}^{u,v} \mid s_{i,i}^{u,v} = V_k(d_i^{(u)}, d_i^{(v)})\} \\ u &= 1,...,N_i \text{ and } v = u+1,...,N_i \\ i &= 1,...,M \end{aligned} \quad (1)$$

where V_k is the k-th comparator for all available $k=1,...,K$ comparators. This will generate $P(N_i,2)=N_i(N_i-1)$ elements in the set S_{ii} where there are N_i-1 mated comparisons per each sample of subject i.

Note that it is assumed that the generation of comparison scores are not symmetric, that is $V_k(d_i^{(u)}, d_j^{(v)}) \neq V_k(d_j^{(v)}, d_i^{(u)})$. If there is evidence of symmetric comparison scores, then the procedure could be modified to only use half of the comparison scores.

2) Generate the set of all non-mated comparison scores using the k-th comparator by comparing samples from person i with samples from all $j=1,...,M$ and $i \neq j$ other persons,

$$\begin{aligned} S_{ij} &= \{s_{i,j}^{u,v} \mid s_{i,j}^{u,v} = V_k(d_i^{(u)}, d_j^{(v)})\} \\ u &= 1,...,N_i \text{ and } v = 1,...,N_j \\ i &= 1,...,M \text{ and } j = 1,...,M \text{ and } i \neq j \end{aligned} \quad (2)$$

The result is $\sum_{j=1, j\neq i}^{M} N_j$ non-mate comparison scores per sample $d_i^{(u)}$.

3) Insert (i,u) into set T if its mated comparison scores is larger than all its non-mated comparison scores, i.e. $s_{i,i}^{u,v} > s_{i,j}^{u,w} \quad \forall j \neq i, v \neq u, w$. This is a rank 1 condition.

4) compute the target utility for sample $d_i^{(u)}$ as

$$utility_i^u = \frac{m_{i,u}^{mated} - m_{i,u}^{non-mated}}{\sigma_{i,u}^{mated} + \sigma_{i,u}^{non-mated}} \quad (3)$$

where $m_{i,u}^{mated}$ is the mean of sample $d_i^{(u)}$'s mated comparison scores computed as:

$$m_{i,u}^{mated} = \frac{\sum_{\substack{v=1 \\ v \neq u}}^{N_i} s_{i,i}^{u,v}}{N_i - 1} \quad (4)$$

and $m_{i,u}^{non-mated}$ is the mean of sample $d_i^{(u)}$'s non-mated comparison scores computed as:

$$m_{i,u}^{non-mated} = \frac{\sum_{\substack{j=1 \\ j \neq i}}^{M} \sum_{v=1}^{N_j} s_{i,j}^{u,v}}{\sum_{\substack{j=1 \\ j \neq i}}^{M} N_j} \quad (5)$$

similarly, $\sigma_{i,u}^{mated}$ is the standard deviation of sample $d_i^{(u)}$'s mated comparison scores computed as:

$$\sigma_{i,u}^{mated} = \sqrt{\frac{\sum_{\substack{v=1 \\ v \neq u}}^{N_i} (s_{i,i}^{u,v} - m_{i,u}^{mated})^2}{N_i - 1}} \quad (6)$$

and $\sigma_{i,u}^{non-mated}$ is the standard deviation of sample $d_i^{(u)}$'s non-mated comparison scores computed as:

$$\sigma_{i,u}^{non-mated} = \sqrt{\frac{\sum_{\substack{j=1 \\ j \neq i}}^{M} \sum_{v=1}^{N_j} (s_{i,j}^{u,v} - m_{i,u}^{non-mated})^2}{\sum_{\substack{j=1 \\ j \neq i}}^{M} N_j}} \quad (7)$$

Once all target utilities ($utility_i^u \ \forall u = 1,...,N_i$ and $\forall i = 1,..,M$) have been computed:

a) Compute two empirical cumulative distribution functions: One for the top-ranked mated comparison scores of set T

$$C(z) = \frac{\left|\left\{utility_i^u : (i,u) \in \mathrm{T}, utility_i^u \le z\right\}\right|}{\left|\left\{utility_i^u : (i,u) \in \mathrm{T}, utility_i^u < \infty\right\}\right|} \qquad (8)$$

and another for those not in that set.

$$W(z) = \frac{\left|\left\{utility_i^u : (i,u) \notin \mathrm{T}, utility_i^u \le z\right\}\right|}{\left|\left\{utility_i^u : (i,u) \notin \mathrm{T}, utility_i^u < \infty\right\}\right|} \qquad (9)$$

These cumulative distribution functions are plotted in Figure A.1 for live-scan images of the right-index fingers of 6,000 subjects and scores of a commercial fingerprint comparator where $N_i = 2$ for all $i = 1,...,6000$.

b) Select target quality resolution (L). That is the number of quality levels for the target quality and therefore target quality scores will be $q = 1,...,L$ where 1 is the lowest and L is the highest quality score. L could be any integer between 2 and 100, for example 5. Note that the larger L, the more samples (larger M) are needed for accurate quality assignments.

c) Bin sample target utilities into L bins based on quantiles of the target utility distributions *C(.)* and *W(.)* in equations 8 and 9. Bin boundaries could be chosen to be cumulative distribution functions (i.e. *C(.)* and *W(.)*) turning points. One strategy, for $L = 5$, is shown in Table A.1 in which W^{-1} and C^{-1} are the quantile functions, and $C^{-1}(0)$ and $C^{-1}(1)$ denote the empirical minima and maxima, respectively (same for $W^{-1}(0)$ and $W^{-1}(1)$), *x* and *y* are appropriate percentile points selected based on the shape of $C(.)$. Bin boundaries for *x=0.25* and *y=0.75* are shown in Figure A.1.

d) Sample $d_i^{(u)}$ is assigned target quality $q_i^{(u)}$ corresponding to the bin of its target utility from equation (3).

The procedure is repeated for all samples $d_i^{(u)} \ \ u = 1,...,N_i$ and $i = 1,...,M$ and all K comparators. Since one sample will have mated comparison scores and a non-mated comparison score distribution different from another sample, different samples of the same subject may have different target utilities and therefore different target quality scores.

II. Aggregate result of binning for K comparators. Choice of aggregate function depends on if and to what degree generalizability of target quality scores is desirable. Below is a list of some options:

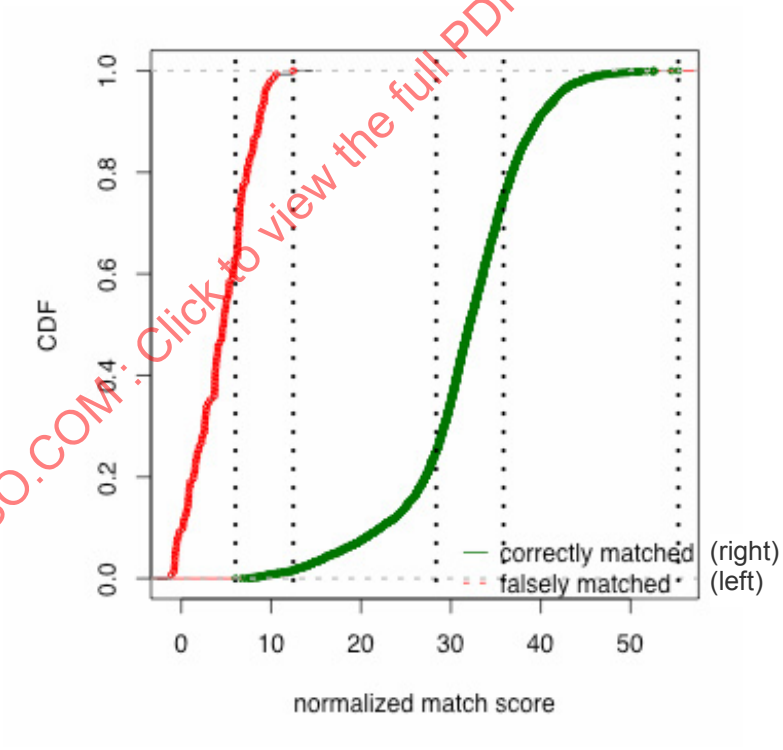
a) unanimity: Samples with identical quality assignments from all K comparators become members of the Quality Reference Dataset. Those without unanimity can be discarded.

b) median or other specified percentile point: Samples with identical quality assignment from more than *X* percent of K comparators become members of the Quality Reference Dataset. The rest can be discarded. Note that *X=100* is the unanimity, and *X=50* is the majority vote rule.

c) arithmetic mean: Final target quality score of each sample will be the arithmetic mean of its quality assignment from all K comparators.

Table A.1 — Binning target utilities

Bin	Range of target utilities
1	$\{z_i : -\infty < z_i < C^{-1}(0.01)\}$
2	$\{z_i : C^{-1}(0.01) \le z_i < W^{-1}(1)\}$
3	$\{z_i : W^{-1}(1) \le z_i < C^{-1}(x)\}$
4	$\{z_i : C^{-1}(x) \le z_i < C^{-1}(y)\}$
5	$\{z_i : C^{-1}(y) \le z_i\}$



NOTE The vertical lines are one possible way of binning normalized comparison scores.

Figure A.1 — Empirical and cumulative distribution functions for the top-ranked genuine scores and for the impostor scores

A.4 Size of quality score normalization corpus

The QSN corpus should be sufficiently large to allow for it to be used for both training and testing of the various proprietary quality scoring algorithms in order to normalize the output to conform to the categories provided by the QSN corpus in this document. The sample size, N, for a controlled trial can be estimated using:

$$N > 32\frac{s^2}{d^2} \tag{8}$$

where s is typical error (noise) and d is smallest worthwhile effect (signal).

Since there is a minimum of four categories, it can be assumed that *d*=1 as the increment for the category is only one. The most typical error is the wrong allocation of the quality into its closest category instead of the correct category, which gives *s*=2. Thus based on equation (8), the minimum required size for the corpus per category is estimated at 128 or 512 per QSN corpus. Half of the corpus could be used for training and the other half for testing. A two-fold cross-validation approach could be used to compute the overall error rate whereby the training and testing sets are interchanged and the average error rate is computed based on the outcome of each test set as has been suggested in the statistical and machine learning literatures.