International Standard

ISO/IEC 29794-1

Third edition
2024-05

Information technology — Biometric sample quality —

Part 1: Framework

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

Partie 1: Cadre

Reference number
ISO/IEC 29794-1:2024(en)

COPYRIGHT PROTECTED DOCUMENT

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

Published in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives or www.iec.ch/members_experts/refdocs).

ISO and IEC draw attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO and IEC take no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO and IEC had not received notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents and https://patents.iec.ch. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

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

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

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

This third edition cancels and replaces the second edition (ISO/IEC 29794-1:2016), which has been technically revised.

The main changes are as follows:

— the definitions of “quality”, “quality score”, and “utility” have been aligned with those in ISO/IEC 2382-37:2022;

— methods for evaluating the efficacy of quality assessment algorithms have been added;

— ASN.1 encoding as defined in ISO/IEC 39794-1 is supported.

A list of all parts in the ISO/IEC 29794 series can be found on the ISO and IEC websites.

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

Introduction

Quality measures are useful for several applications in the field of biometrics. While ISO/IEC 19784-1 specifies a structure and gives guidelines for quality score categorization, this document defines and specifies methodologies for objective and quantitative quality score expression, interpretation and interchange.

This document establishes a framework that facilitates the use of biometric sample quality assessment and scoring tools. The tools are intended to encourage innovation and performance improvements in, and interoperability of, biometric systems generally. The ISO/IEC 29794 series presents several biometric sample quality assessment and scoring tools, the use of which is generally optional but can be determined as mandatory by particular application profiles or specific implementations. The ISO/IEC 29794 series is prepared to accommodate additional parts that address the biometric modes specified by the ISO/IEC 19794 series and the ISO/IEC 39794 series, with part numbers and titles aligning appropriately. However, as this document is intended for use by all biometric modes, a mode does not necessarily need a mode-specific part to make use of quality scores.

Several applications can benefit from the use of biometric sample quality measures. An example is the use of real-time quality feedback as part of the biometric capture process to improve the operational efficiency and performance of a biometric system. Other examples include data fusion for which multiple samples or references are available in the comparison process, either from a single or multiple biometric mode, and hardening systems against presentation attacks using or targeting low quality biometric samples. The association of quality measures with biometric samples is an important component of quality measure standardization. Quality fields as specified in Clause 7 are included in biometric data interchange formats. If a CBEFF (Common Biometric Exchange Formats Framework) header is present, then CBEFF_BDB_quality may additionally be used to express quality measures. Useful analyses can be performed using quality measures along with other data to improve the performance of a biometric system. For example, correlating quality measures 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

This document establishes the following items for any or all biometric sample types as necessary:

— terms and definitions that are useful in the specification and use of quality measures;

— purpose and interpretation of biometric quality scores;

— motivation for developing biometric sample datasets for the purpose of quality score normalization;

— format for exchange of quality assessment algorithm results;

— methods for aggregation of quality scores;

— methods for evaluating the efficiency of quality assessment algorithms.

The following are outside the scope of this document:

— specification of minimum requirements for sample, module, or system quality scores;

— standardization of quality assessment algorithms;

— assessment of utility of biometric samples or references for human examiners.

2 Normative references

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

ISO/IEC 39794-1, *Information technology — Extensible biometric data interchange formats — Part 1: Framework*

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

ISO/IEC 19785-2, *Information technology — Common Biometric Exchange Formats Framework — Part 2: Biometric registration authority*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 2382-37, ISO/IEC 39794-1 and the following apply.

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

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

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

3.1
acquisition fidelity
fidelity (3.8) of a biometric sample attributed to the acquisition process

3.2
biometric character
set of attributes associated with a biometric characteristic that cannot be controlled during the biometric acquisition process

EXAMPLE Scars, number of minutiae, blepharoptosis (droopy eyelid)

[SOURCE: ISO/IEC 2382-37:2022, 37.09.15, modified — Note 1 to entry has been removed.]

3.3
biometric utility
degree to which a biometric sample supports biometric recognition *performance* (3.11)

Note 1 to entry: The *biometric character* (3.2) of the sample source, the *fidelity* (3.8) of the processed biometric samples and the conformance of the biometric sample presentation contribute to, or similarly detract from, the utility of the biometric sample.

Note 2 to entry: Performance measures such as false match rate, false non-match rate, failure-to-enrol rate, and failure-to-acquire rate are an indication of biometric utility.

[SOURCE: ISO/IEC 2382-37:2022, 37.09.16, modified — "character" has been changed to "biometric character" in Note 1 to entry.]

3.4
environment
physical surroundings and conditions in which the biometric capture takes place

Note 1 to entry: The conditions include the factors such as lighting and temperature, level of enrolee cooperation, and the skill of the operator, if one is involved in the capture process.

3.5
false non-match error versus discard characteristic
FNM-EDC
method to evaluate the efficacy of *quality assessment algorithms* (3.13) by quantifying how efficiently discarding samples with low *quality scores* (3.16) results in an improved (i.e. reduced) false non-match rate

Note 1 to entry: The false non-match error versus discard characteristic is a graphical presentation of the *performance* (3.11) of quality assessment algorithms, plotting the dependence of the false non-match rate at a fixed comparison decision threshold on the percentage of low-quality reference and probe samples discarded.

3.6
false match error versus discard characteristic
FM-EDC
method to evaluate the efficacy of *quality assessment algorithms* (3.13) by quantifying how efficiently discarding samples with low *quality scores* (3.16) results in an improved (i.e. reduced) false match rate

Note 1 to entry: The false match error versus discard characteristic is a graphical presentation of the *performance* (3.11) of quality assessment algorithms, plotting the dependence of the false match rate at a fixed comparison decision threshold on the percentage of low-quality reference and probe samples discarded.

3.7
extraction fidelity
component of the *fidelity* (3.8) of a sample attributed to the biometric feature extraction process

3.8
fidelity
degree to which a biometric sample is representative of its source biometric characteristic

Note 1 to entry: The fidelity of a sample comprises components attributable to one or more of the processing steps: acquisition, extraction, signal processing.

3.9
interpretation
process of analysing a *quality score* (3.16) along with other data in order to give that score contextual, relative meaning

3.10
native quality measure
output of a *quality assessment algorithm* (3.13) without constraints on data format and/or value range

3.11
performance
assessment of false match rate, false non-match rate, failure-to-enrol rate, failure-to-acquire rate, processing time or throughput rates of a biometric system

3.12
quality
degree to which a biometric sample meets the specified requirements for its targeted application

[SOURCE: ISO/IEC 2382-37:2022, 37.09.14]

3.13
quality assessment algorithm
quality algorithm
algorithm to calculate a *quality measure* (3.15)

Note 1 to entry: The ISO/IEC 19785 series uses the term "quality algorithm".

3.14
quality component
measurement on the biometric sample that may contribute to the computation of a unified *quality score* (3.16)

Note 1 to entry: Features expressing quality components are defined in the modality-specific parts of the ISO/IEC 29794 series.

3.15
quality measure
quality score (3.16) or *quality component* (3.14)

3.16
quality score
quantitative value of the fitness of a biometric sample to accomplish or fulfil the comparison decision

[SOURCE: ISO/IEC 2382-37:2022, 37.09.13]

3.17
quality score normalization
rescaling of *quality scores* (3.16) to improve consistency in scale and *interpretation* (3.9)

3.18
quality score normalization dataset
QSND
dataset of biometric samples annotated with *quality scores* (3.16) for use in *quality score normalization* (3.17)

Note 1 to entry: Target quality scores may be assigned based on *performance* (3.11) outcomes using the sample in question or may be based on quality factors recorded in the acquisition of the dataset.

3.19
quality score percentile rank
QSPR
percentile rank of *quality scores* (3.16) of biometric samples in an identified control dataset that are less than the specified quality score

Note 1 to entry: See *QSND* (3.18).

3.20
raw quality score
quality score (3.16) that has not been *interpreted* (3.9), either by the creator or recipient of the score, and alone potentially does not intrinsically provide contextual information

4 Abbreviated terms

BDB	biometric data block
CBEFF	common biometric exchange formats framework (ISO/IEC 19785)
CDF	cumulative distribution function
DET	detection error trade-off
FERET	facial image database developed by the U.S. government in the 1990s
FMR	false match rate
FNMR	false non-match rate
QAID	quality assessment algorithm identifier
QSND	quality score normalization dataset
QSPR	quality score percentile rank
QVID	quality assessment algorithm vendor identifier

5 Conformance

A biometric sample quality block shall be considered conformant to this document if its structure and data values conform to the formatting requirements of Clause 7.

The semantic conformance testing will be handled in the modality-specific parts of the ISO/IEC 29794 series, where, for example, conformance test sets (a set of biometric samples representing the entire variety of quality from poor to good) and associated quality scores to be obtained with the reference implementation are given.

6 Biometric sample quality criteria

6.1 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 among the different aspects of quality, illustrated in Figure 1. The quality of a biometric sample depends on character and fidelity. Figure 2 illustrates the relationship between quality (character, fidelity and utility) and system performance. The utility of a biometric sample reflects the impact of this sample on biometric recognition performance.

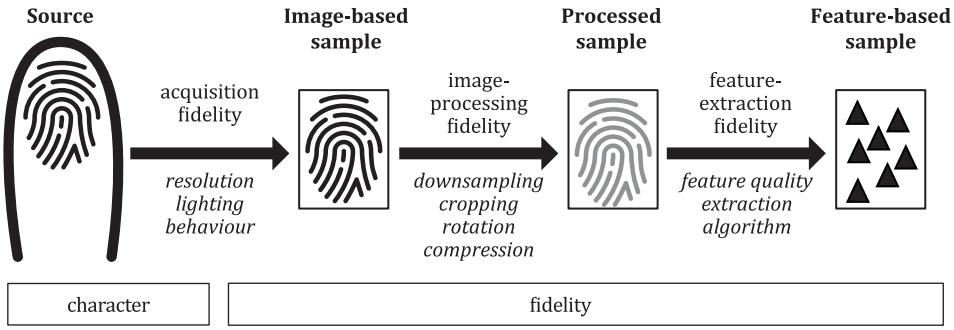


Figure 1 — Quality reference model illustration

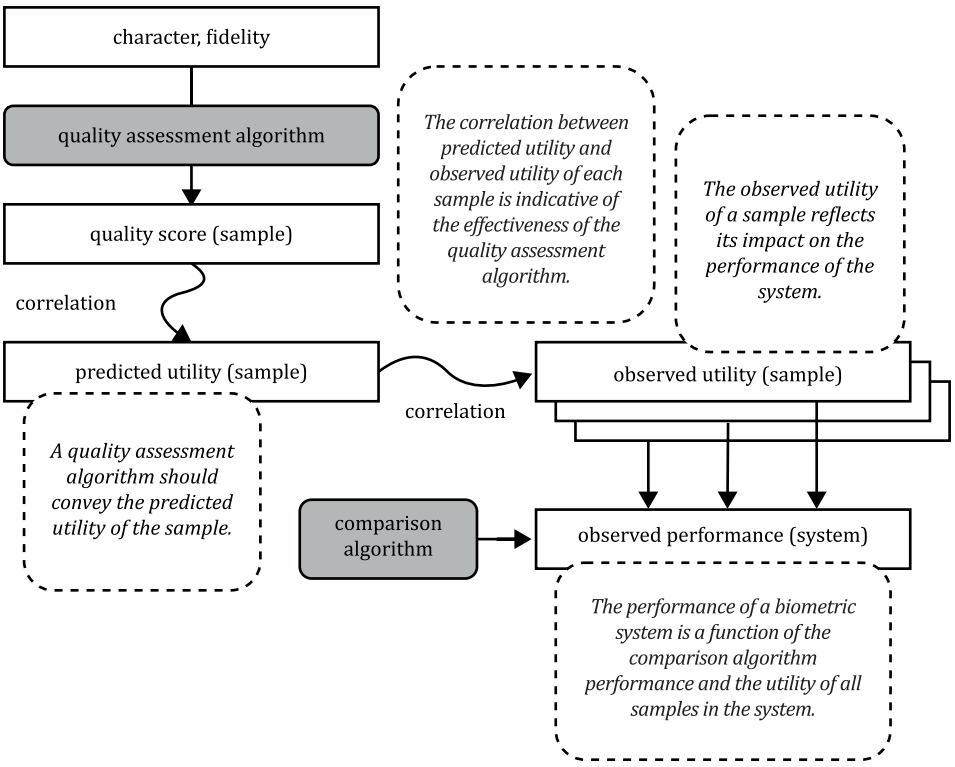


Figure 2 — Relationship between quality and system performance

6.2 Quality aspects: 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 subjectively reflect the following.

a) Character of a sample — An expression of quality based on the inherent properties of the biometric characteristic from which the biometric sample is derived. For example, worn friction ridges have poor character and blepharoptosis (droopy eyelid) causes poor iris character.

b) Fidelity of a sample to the biometric characteristic from which it is derived — An expression of quality based on fidelity reflects how accurately the sample represents its biometric characteristic. Sample fidelity is comprised of fidelity components contributed by different processes.

c) 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 or reference as well as the details of the specific biometric system of which performance is being evaluated. This implies that utility is not necessarily a universal attribute of a sample consistent across all systems.

Utility-based quality is intended to be more predictive of system performance (e.g. in terms of false match rate, false non-match rate, failure to enrol rate, and failure to acquire rate) than measures of quality based on character or fidelity alone. See Table 1 for more information.

The term “quality” is not solely attributable to the characteristics of the capture device, such as sampling rate, transfer function, directionality, sensitivity, dynamic range and bit depth, image resolution, pixel density, dimensions in pixels, or grey scale/colour bit depth, although such factors can affect sample utility and can contribute to the overall quality score.

The character and utility of an acquired sample depend on the features generated by a feature extraction subsystem. For instance, the same finger image can 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.

As avoidance of demographic differentials in performance across populations is vitally important to all applications of biometrics, quality measures should not be based on performance measures that correlate with age, ethnicity, gender, sex, religion or recognized disabilities. For this reason, quality measures should be described to the extent possible so that metrics with a potential demographic differential can be recognized.

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 can 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. Recapture or signal enhancement techniques can improve utility.	Samples with high character and high fidelity indicate capture of a useful sample. High utility is expected.

6.3 Use cases of data quality measures

6.3.1 General

This document restricts the definition of "utility" to the performance of automated systems for the recognition of individuals based on their biological and behavioural characteristics. Assessment of utility of biometric samples and references for human examination or forensic applications is beyond the scope of this document.

6.3.2 Real-time quality assessment

Real-time quality assessment of a biometric sample and the resulting quality measures can be used by an operator, by an automated system, or by a biometric data subject to help improve the average quality of captured biometric samples. This feedback can be used in manual or automated decision-making to determine whether another capture attempt is needed, or whether a sample should be accepted or discarded and not be used for enrolment or comparison. This provides the opportunity for overall system performance to be improved by assisting an operator or augmenting an automated quality control system in the context of decisions as to whether to accept or retain the sample, discard the sample, reattempt a capture, or declare a failure to acquire or failure to enrol. Quality measures can be retained for later use, for example, for determining whether an enrolment sample should be replaced when the next sample is captured.

6.3.3 Use in different applications

An acquired biometric sample can be used in multiple applications involving several different feature extraction and comparison algorithms. These applications and algorithms can be unknown at the time the

sample is acquired and its quality is assessed. As far as possible, the assessed quality of the sample should be broadly predictive of utility across uses and biometric system algorithms.

One challenge in establishing a universal quality standard is in defining a measure that is sufficiently adaptable for use with different comparison algorithms across applications with varying utility metrics. Thus, a quality assessment algorithm will be likely to produce measures of predicted utility for only a limited number of biometric systems. It can be useful to compute and apply multiple quality scores to improve predictability of various failure modes.

A second challenge is that comparison algorithms produce scores from the comparison of a probe to a reference and are influenced by the quality of each. If the reference exists in the form of an aggregated or averaged sample, or is a model, it will not necessarily be possible to assign a quality score to the reference.

A third challenge is that reference databases are generally curated or created under various policy-driven constraints, either implicit or explicit. For example, blank fingerprint images or empty minutiae files are generally removed from a fingerprint database. Facial images used as references can be limited to those meeting the requirements of ISO/IEC 39794-5:2019, Clause D.1. Speaker recognition models can have been developed using a particular audio acquisition channel. Similarity scores resulting from comparisons of probes with a reference will be affected by the extent to which the probe collection mimics the reference collection and curation policies.

Therefore, in developing a quality assessment algorithm, it is necessary to state the assumptions of the reference creation and curation process as completely and clearly as possible. For example, one face image quality assessment algorithm can be developed for full-frontal reference face images that conform to ISO/IEC 39794-5:2019, Clause D.1, whereas another one can be developed for in-the-wild face images.

It is useful for algorithm-specific quality scores to be interpretable within the context of the capture device producing the original biometric sample and the application to which they are being applied. The ability to interpret quality scores within the context of both their generation and application is particularly important in the setting of comparison decision thresholds (or recognition threshold).

6.3.4 Use as a survey statistic

Quality scores can be used to monitor operational conditions and processes.

EXAMPLE 1 Aggregated quality scores can be compared with pre-set limits or monitored against an operational requirement. See Annex C for procedures for aggregation.

EXAMPLE 2 If quality scores are generated from biometric samples collected at many sites, or over different time periods, then they can be used to identify anomalous operation.

EXAMPLE 3 If face image quality is computed at the licence issuance desks at a Department of Motor Vehicles, then a ranked list of aggregated quality scores can be used to identify desks that exhibit a lower-than-average quality, or to monitor trends over weeks or months.

6.3.5 Accumulation of relevant statistics

Reliable quality scores can 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 of whether a higher quality sample is likely if another capture is attempted.

6.3.6 Sample-based reference database improvement

The association of quality measures with a sample that is to be entered into a reference database is important for the maintenance and improvement of reference database utility. The tracking of sample quality measures can lead to detection of deterioration of operator performance, environmental conditions, or biometric sample capture device performance. Tracking of the sample quality measures should be an important part of a biometric system’s operating procedures. Improvement of the sample reference database can be made by replacement or augmentation to make use of the highest quality biometric sample. Typically, replacement decisions are linked to the comparator performance of the system processing the data.

6.3.7 Quality-based conditional processing

Biometric samples can be processed differently based on quality measures. In particular, poor-quality biometric samples can be processed using different algorithms or thresholds from those used for high-quality biometric samples.

Quality scores should not be used for the detection of presentation attacks. Manipulation of signals (e.g. by adding noise) can affect quality while generating a threat vector.

6.3.8 Quality-directed fusion

When applying multimodal or multi-sample biometric fusion, the relative qualities of samples can be used to direct or augment a fusion process.

In a multi-instance system, the weights for each contributing channel can be determined based on the quality of the biometric sample. For instance, in a ten-print fingerprint recognition system, less weight can be expected for the little finger.

6.3.9 Interchange of quality measures by disparate systems

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

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

6.3.10 Workload reduction with quality scores

In a large-scale biometric system, a nearest quality score-based intelligent search for reduction of the computational workload in biometric identification can reduce the transaction time.[9] More precisely, the variability of quality scores exhibited on different biometric characteristic types (face, iris and fingerprint) can be turned into an advantage for rapid indexing. Depending on the size and properties of the database, the search space can be reduced significantly for each biometric characteristic, depending on the variation in terms of sample quality.[9]

Quality-based processing would improve the efficiency but can adversely affect the overall recognition accuracy due to possible failures of the quality assessment algorithm used.

6.3.11 Selection of the best of a series of biometric samples

Given a series of biometric samples of a data subject, quality scores can be used in the selection of the best sample. This operation is useful when a receiving system expects exactly one sample and the sending system is required to determine which of several collected samples to transmit.

7 Data interchange format field definition

7.1 Abstract description

7.1.1 Overview

Figure 3 illustrates the structure of a quality block. The data structure is designed for the interchange of values of quality measures. The other parts of the ISO/IEC 29794 series may use the data structure for encoding mode-specific quality components (e.g. number of minutiae in a fingerprint image, pose-angle of a captured face).

If no quality scoring is attempted, then there shall be no quality block present. If there is more than one quality measure for a biometric sample, then a sequence of quality blocks shall be used.

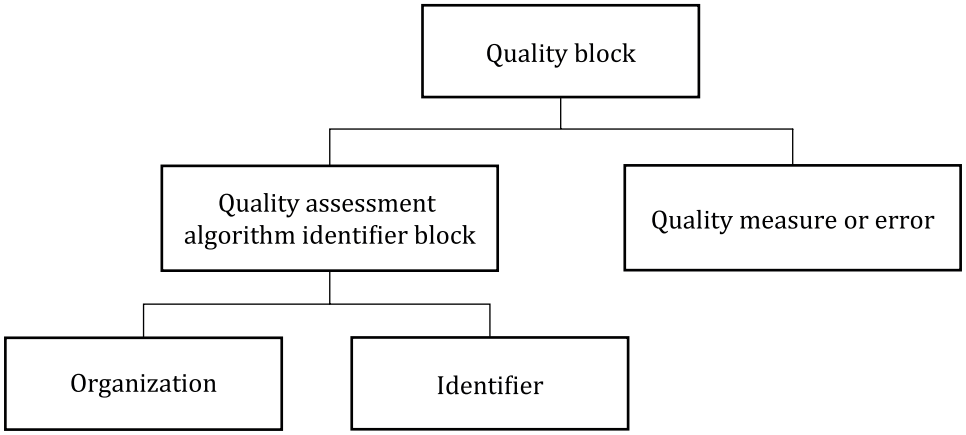


Figure 3 — Structure of a quality block

7.1.2 Quality assessment algorithm identifier block

Abstract values: Sequence of two integers 1 to 65 535

Contents: This data element shall identify the quality assessment algorithm used. It shall consist of two elements:

— the quality assessment algorithm vendor identifier (QVID); and

— the quality assessment algorithm identifier (QAID).

The QVID shall be one of the biometric organization identifiers registered in accordance with ISO/IEC 19785-2. The QAID shall be one of the quality assessment algorithm identifiers associated with the given QVID. Different versions of a quality assessment algorithm that yield different results shall be assigned different QAIDs to allow for unique identification.

NOTE 1 ISO/IEC 19785-1:2020, 7.1.6 states that registration of biometric product identifiers is optional.

NOTE 2 It is indispensable to enable the recipient of biometric data to differentiate between quality measures generated by different quality assessment algorithms and to adjust for any differences in processing or analysis as necessary. The combination of QVID and QAID is a solution that can be implemented quickly but only partially achieves the goals of quality score standardization. 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).

NOTE 3 The other parts of the ISO/IEC 29794 series specify standardized computation methods for the quality measures defined in that part of the ISO/IEC 29794 series.

7.1.3 Quality measure (quality score or quality component) or error

Abstract values: Integer between 0 and 100 or failureToAssess

Contents: Quality measures shall be embedded in the quality block as an integer between 0 to 100. If the output of a quality assessment algorithm is a floating-point number or outside the range from 0 to 100, it shall be scaled to the range from 0 to 100 and rounded to the nearest integer for embedding in the quality block. The abstract value failureToAssess shall indicate that the quality assessment algorithm has failed.

Quality scores enable discrimination between distinct levels of performance. A quality score shall predict performance metrics such as false match and false non-match rates when comparisons are made to references developed under stated collection policies.
EXAMPLE 1 A particular face image quality assessment algorithm can produce quality scores predicting performance against full-frontal reference face images conformant to ISO/IEC 39794-5:2019, Clause D.1.

A quality score represents the entire biometric sample quality in a holistic manner.

A quality score may be a composite of several quality components.
EXAMPLE 2 The quality score of a fingerprint image reflects the print's clarity, uniformity of ridges and valleys, and the number of correctly identified minutia, among other components.

Higher quality score values imply higher biometric utility. Unlike higher values of quality scores, higher values of quality components do not necessarily imply higher biometric utility.
To be predictive of performance, a quality score may model known failure modes/ sensitivities of a biometric comparator and image or signal processing algorithm. 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. fingerprint comparison algorithms based on minutia data). If the biometric system utilizes subsystems from multiple vendors, the quality score should reflect the aspects of performance important for each subsystem used.
NOTE 1 As it is challenging to find a single quality measure that is universal, not vendor-specific and yet adequately indicates performance, it can be useful to apply more than one quality assessment algorithm.

Any time a biometric sample undergoes a transformation (e.g. downsampling or further compression), the quality of the transformed sample should be reassessed and associated with the transformed sample.
EXAMPLE 3 Throughout an identity management system, a biometric sample can be stored in multiple formats (e.g. high-resolution finger image stored centrally and a minutiae-based representation stored on a smart card).

The native quality measure may be scaled using Formula (1):

$$Q_{s,i} = \min\left(\max\left(0{,}100 \frac{Q_{n,i} - \min(Q_n)}{\max(Q_n) - \min(Q_n)} \right), 100 \right) \quad (1)$$

where

$Q_{s,i}$ is the scaled quality measure for biometric sample i;

$Q_{n,i}$ is the native quality measure for biometric sample i;

$\min(Q_n)$ is the minimum value of the native quality measure; and

$\max(Q_n)$ is the maximum value for the native quality measure.

$\min(Q_n)$ and $\max(Q_n)$ may be computed empirically.
NOTE 2 The linear nature of min-max function [see Formula (1)] allows for accurately estimating the minimum and maximum observable values.

Another option for scaling a native quality measure into [0, 100] is a sigmoid function, as shown in Formula (2):

$$Q_{s,i} = \frac{100}{1 + e^{\frac{Q_{n,0} - Q_{n,i}}{w}}} \quad (2)$$

where

$Q_{s,i}$ is the scaled quality measure for biometric sample i;

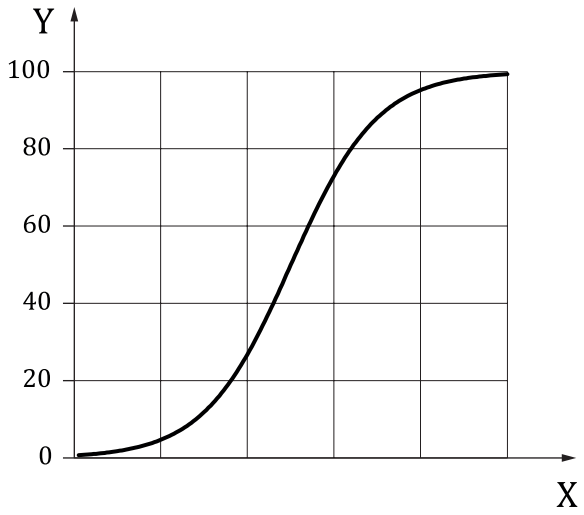
$Q_{n,i}$ is the native quality measure for biometric sample i;

$Q_{n,0}$ indicates the inflection point; and

w is the slope of the sigmoid function.

Parameters $Q_{n,0}$ and w shall be selected by the developer of the quality assessment algorithm or biometric system operators. The sigmoid function output values ($Q_{s,i}$) plotted versus the native quality measure $Q_{n,i}$ is shown in Figure 4.

NOTE 3 The sigmoid function is used for non-linear normalization of continuous features. Instead of using a linear normalization a non-linear normalization allows to focus on those quality values that have higher utility than the values outside the focus area.



Key

X $Q_{n,i}$ (native quality measure for biometric sample i)

Y $Q_{s,i}$ (scaled quality measure for biometric sample i)

Figure 4 — Sigmoid function

7.2 XML encoding

ISO/IEC 39794-1 defines an XML encoding of quality blocks. See Annex A for an example.

7.3 Tagged binary encoding

ISO/IEC 39794-1 defines the abstract syntax of quality blocks in ASN.1. The tagged binary encoding of a biometric data block is obtained by application of the ASN.1 Distinguished Encoding Rules (ISO/IEC 8825-1) to the ASN.1 module describing the data block. See Annex A for an example.

8 Exchange of quality assessment algorithm results

Quality assessment algorithm vendors should be able to offer results of their quality assessment algorithms in a standardized way to the biometric community. On the other hand, consumers of ISO/IEC 19794 series and ISO/IEC 39794 series biometric data blocks can retrieve and process this information effectively to assess the value of the output of this quality assessment algorithm to their implementation. This approach has the following benefits.

— Both quality assessment algorithm vendors and consumers can gain value from technical improvement, which is necessary in the starting phase of widespread 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 assessment algorithm vendors (who carry out the evaluation).

— Over time, standardized test sets will evolve, for the following reasons:

  — it is in the interest of the quality assessment algorithm vendor to use a reporting test set(s), that is of use for many customers, and

  — the need for new test sets will diminish over time and the use of new tests will be critically reviewed by the biometric community.

— Evolution of test sets will facilitate the development of QSND.

For the exchange, the following items shall be provided:

a) quality assessment algorithm vendor ID;

b) quality assessment algorithm ID;

c) minimum and maximum theoretical output value of the algorithm;

d) unique name of the test set used (e.g. in form of "FERET-Greyscale" in the case of face recognition);

e) list of samples that have been processed.

Anyone can publish new test sets (biometric samples and a naming scheme).

A self-describing language like XML should be used for the description of the data sets as well as for the evaluation results. The evaluation results can 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 Quality score normalization

Normalization of quality score data is the process by which quality score data is processed by its recipient to give the scores local context and meaning, for example, making quality scores from different algorithms have similar meaning.

A raw quality score is assigned to a biometric sample by a particular quality assessment algorithm. To interpret the raw score, the recipient of a score shall have some contextual information. This information can be provided in the following ways.

a) Extrinsically, in the form of metadata or offline data (e.g. standard) that instruct the recipient on interpretation of the score. As a quality score is accompanied by the identifier of the algorithm used to generate it (i.e. QAID), recipient software can be configured to use vendor-supplied data (e.g. suggested thresholds) to best process the sample. The algorithm can alternatively be used to perform analysis to fully optimize the interpretation of the scores given the local application and data. By identifying the

algorithm, scores created by different algorithms can be differentiated so that, for example, different thresholds could be applied to the sample depending on the source of the quality score.

b) Intrinsically, in the form of a normalized quality score. Normalization of quality score data provides contextual information about the score.

QAID enables vendor-specific scaling, such that the 0 to 100 scale correlates to another scale reflecting the above. For example, the recipient of a file would be encouraged to analyse the relationship between quality scores and the false match rate and false non-match rate of the samples processed by their comparator. The results could be used, for example, to specify an operating quality threshold for sample acquisition or discarding. This method provides the recipient with 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 a quality score normalization dataset (QSND) 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 assessment algorithm "X" would consider biometric sample "Y" to have a quality percentile rank "Z" if compared to the data in the QSND. The translation of raw quality scores to percentile rank scores is achieved by running a standardized corpus of samples through a given quality assessment algorithm and pairing all possible raw score outcomes to percentile rank scores.

10 Pairwise quality

Comparison scores result from the comparison of a probe to a reference sample, both of which may have quality scores attached. In some applications, there are no assumptions regarding the conformance of either the probe or the reference to any collection best practices or requirements. Utility, as described in 6.2, should not be considered solely as a function of the conformance of either the probe or the reference to any collection requirements. The issue of assigning a single quality score that is predictive of performance and is reflective of the comparison score computed from the two biometrics samples involved in the comparison should be addressed. The assignment of pair-wise quality can be useful for applications using dynamic recognition thresholds, comparison score level fusion, or evaluation of quality assessment algorithms as detailed in Clause 11.

Consider a probe sample (superscript 1) from data subject *i*, and a reference (superscript 2) from data subject *j*, where the probe and reference samples are assumed to be representations from the same biometric characteristic and mode. If *i* and *j* are different data subjects, and the samples represent the same biometric instance (e.g. the same index finger), the comparison is "non-mated." If *i* and *j* are the same data subject, and the samples represent the same biometric instance (e.g. the same index finger), the comparison is "mated". The quality of the probe sample is $q_i^{(1)}$ and the quality of the reference sample is $q_j^{(2)}$. Comparison of the probe sample to the reference sample results in a comparison score s_{ij}, which must be associated with a single quality measure $Q(s_{ij})$. The function for assigning the single quality measure will be called F, such that Formula (3) applies:

$$Q(s_{ij}) \cong F(q_i^{(1)}, q_j^{(2)}) \tag{3}$$

The choice of pairwise quality function F depends on the characteristics of the quality measures and their proposed use. Appropriate choices for F depend upon modality and may include the minimum, mean, difference or other relationship between the probe and reference quality measures. These are discussed within the other parts of the ISO/IEC 29794 series.

Although there is only one comparison score resulting from each probe-to-reference comparison, there can be several probes and references from data subjects *i* and *j*, such that there can be several s_{ij} values for each *i* and *j*. In the interest of notational simplicity, within this document, only one comparison score is considered for each *i* and *j*.

11 Evaluation

11.1 General

This clause lists several methods for evaluating the performance of quality assessment algorithms by evaluating if and how quality scores predict performance of a biometric recognition subsystem. These methods may be applied regardless of whether biometric references meet specified requirements of the targeted application. In the computation of methods in 11.2 through 11.5, if the quality assessment does not apply to the reference samples, only probe quality should be used.

11.2 False non-match error versus discard method

One metric for comparative evaluation of biometric sample quality assessment algorithms is the false non-match error versus discard characteristic curves, introduced as error vs. reject curve with respect to false non-match errors in Reference [8]. The goal is to demonstrate how efficiently discarding of samples with low quality scores results in an improved (i.e. reduced) false non-match rate.

Consider an application where a pair of samples (superscripts 1 and 2) from the same biometric modality of the same data subject, i, with quality scores $q_i^{(1)}$ and $q_i^{(2)}$ are compared to produce a similarity score s_{ii}, where there are N such pairs. As unified quality scores are expected to increase with improving quality, a quality threshold u is introduced to define levels of acceptable quality and define the set $D(u)$ of low-quality entries as shown in Formula (4):

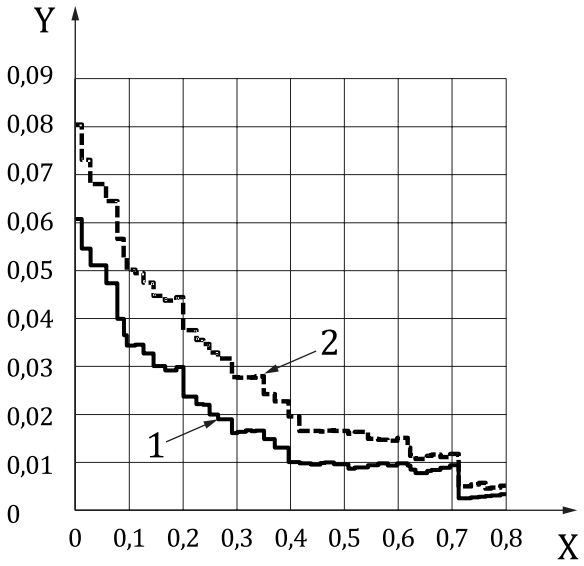
$$D(u)=\left\{ii:F\left(q_i^{(1)},q_i^{(2)}\right)<u\right\} \tag{4}$$

where the pairwise quality function F can be selected to best suit the operational needs and constraints. In many cases, the quality scores can be combined using $F\left(q_i^{(1)},q_i^{(2)}\right)=\min\left(q_i^{(1)},q_i^{(2)}\right)$ in acknowledgement that the lower quality sample drives lower similarity scores, or simply $F\left(q_i^{(1)},q_i^{(2)}\right)=q_i^{(2)}$ if the evaluation or operation has been set up such that the reference sample $q_i^{(1)}$ has high quality by design. Then, the false non-match rate is computed as the fraction of mated similarity scores (e.g. scores from comparison of the same finger of the same subject) below some critical recognition threshold, t, for the sample pairs not in set $D(u)$. The recognition threshold t is fixed, and the quality threshold u is varied across the algorithm-specific quality score domain to show the dependence of false non-match rate on quality threshold at the selected recognition threshold. See Annex D for an example script. The value of false non-match rate as a function of percentage of discarded samples [which depends on quality threshold u – see Formula (4)] for the recognition threshold t can be shown graphically.

If the quality scores are perfectly correlated with the mated similarity scores, setting recognition threshold t to give an overall false non-match rate of x and then discarding x percent of samples with the lowest quality scores will result in false non-match rate of zero. For a good quality assessment algorithm, the false non-match rate should decrease quickly with the fraction discarded.

Comparison scores can be computed from two or more samples. Computing the false non-match error versus discard characteristics based on comparison scores computed from more than two samples is not covered in Formula (4).

Figure 5 shows examples of the false non-match error versus discard characteristics for two recognition thresholds corresponding to false match rates of 0,1 % and 0,01 %, for quality scores computed by NFIQ 2,[4] open-source finger minutiae comparison software[5],[6] and a fingerprint database whose samples were captured by an optical fingerprint capture device.[7]



Key

X discard ratio

Y FNMR

1 initial FMR = 0,1 %

2 initial FMR = 0,01 %

Figure 5 — Examples of false non-match error vs. discard characteristics at two different recognition thresholds corresponding to initial false match rate values of 0,1 % and 0,01 %

11.3 False match error versus discard method

In certain applications where false positives are common and critical, or when a particular recognition algorithm results in high non-mated similarity scores on low quality samples, it can be valuable to assess whether a quality assessment algorithm can predict which samples will result in false matches. Additionally, to examine the full impact of discarding low quality samples on performance, false non-match error versus discard characteristic and false-match error versus discard characteristic should be explored together.

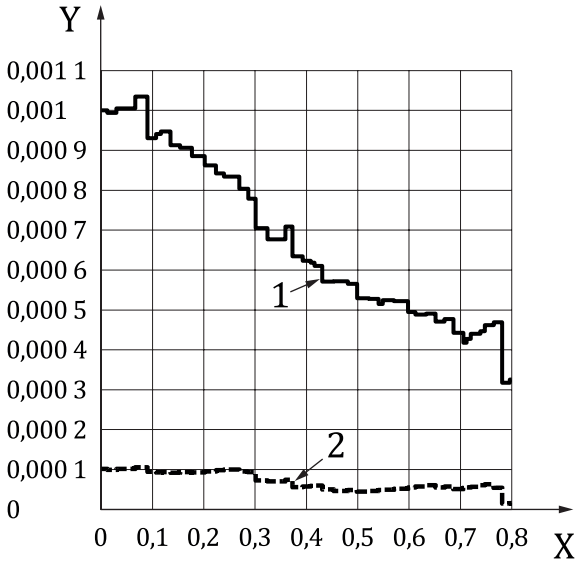
Similarly to the false non-match error versus discard method above, the goal of the false match error versus discard method is to demonstrate how efficiently discarding samples with low quality scores can result in an improved (i.e. reduced) false match rate.

Consider an application where a pair of biometric samples from different data subjects, *i* and *j*, with quality scores q_i and q_j, are compared to produce a similarity score s_{ij}. As quality scores are expected to increase with improving quality, a quality threshold *u* is introduced to define levels of acceptable quality and define the set $D(u)$ of low-quality entries as shown in Formula (5):

$$D(u)=\{ij:F(q_i,q_j)<u\},i\neq j \tag{5}$$

where the quality values are combined using $F(q_i,q_j)=\min(q_i,q_j)$ in acknowledgement that the lower quality sample drives lower comparison scores, or simply $F(q_i,q_j)=q_j$ if the evaluation or operation has been set up such that the reference sample q_i has high quality by design. Then, the false match rate is computed as the fraction of non-mated similarity scores (e.g. comparison of the face images of different data subjects) above some recognition threshold, *t,* for the sample pairs not in set $D(u)$. The recognition threshold *t* is fixed, and the quality threshold *u* is varied across the algorithm-specific quality score domain to show the dependence of false match rate on quality threshold at the selected recognition threshold. See Annex D for an example script. The value of false match rate as a function of quality threshold *u* for the recognition threshold *t* can be shown graphically.

Figure 6 shows examples of the false match error versus discard characteristics for two recognition thresholds corresponding to false match rates of 0,1 % and 0,01 %, for quality scores computed by NFIQ 2,[4] open-source finger minutiae comparison software[5],[6] and a fingerprint database whose samples were captured by an optical fingerprint sensor.[7]



Key

X discard ratio

Y FMR

1 initial FMR = 0,1 %

2 initial FMR = 0,01 %

Figure 6 — Examples of false match error vs. discard characteristics at two different recognition thresholds corresponding to initial false match rate values of 0,1 % and 0,01 %

11.4 DET versus discard method

Another metric for comparative evaluation of biometric sample quality assessment algorithms is the DET versus discard plots. This method can be applied regardless of the type of reference (sample, template, model) being employed. The goal is to demonstrate how efficiently discarding samples with low quality scores results in an improved DET plot. Advantages of this approach are that it simultaneously takes both false non-match and false match errors into consideration and that it does not require a critical recognition threshold, as all possible thresholds are represented.

Consider an application where a sample with quality score q_i is being compared to a reference sample with quality score r_j from either the same or different biometric data subjects, i and j, to produce a similarity score s_{ij}, where there are N pairs from same data subjects and M pairs from different data subjects. Two histograms can be created from these scores: a mated histogram for the s_{ij}, where $i=j$, and a non-mated histogram for the s_{ij}, $i \neq j$. These can be used to create two cumulative distribution functions, MatedCDF(t), and NonMatedCDF(t) for all $s_{ij}<t$, one cumulative curve for each histogram. The DET can be created by plotting the value of false non-match rate =1-MatedCDF(t) against false match rate= NonMatedCDF(t) as t increases across the entire range of comparison-algorithm-dependent values of scores, s_{ij}. A quality threshold u is introduced to define levels of acceptable quality and define the set $D(u)$ of low-quality entries as shown in Formula (6):

$$D(u)=\{q_i<u \text{ OR } r_j<u\}. \quad (6)$$

If the quality assessment does not apply to the reference samples (i.e. r_j is not computed), then $D(u)=\{q_i<u\}$.

For several quality thresholds, u, a family of DET curves is calculated discarding the comparisons that are in the set $D(u)$. For a good quality assessment algorithm, the false non-match rate should decrease over a broad domain of false match rate values with increasing quality threshold u.

11.5 Sample acceptance or discard rate

This subclause defines two related metrics for evaluating the performance of quality assessment algorithms used to make decisions on whether to retain a biometric sample for further processing.

The first metric is an error rate expressing the proportion of biometric samples incorrectly discarded when they would reach correct match decisions by a downstream comparison subsystem. The incorrect sample discard rate (R_{ISD}) is estimated from a set of test outputs comprised of a vector of quality scores, q, a vector of mated comparison scores, s, a recognition threshold, t, and a quality threshold u as shown in Formula (7):

$$R_{\mathrm{ISD}}(u) = \frac{1}{N}\sum_{i}^{N}(1 - H(q_i - u))H(s_i - t) \tag{7}$$

where

N is the number of samples in the vectors;

$H(x)$ is the step-function that is 0 unless x is greater than or equal to 0 when it is 1

Formula (7) gives the proportion of samples with quality scores below the quality threshold u but with comparison scores above a recognition threshold.

The second metric is an error rate expressing the proportion of biometric samples incorrectly retained when they ultimately result in false non-matches by the comparison subsystem. This error rate is called incorrect sample accept rate, R_{ISA} , and is computed as shown in Formula (8):

$$R_{\mathrm{ISA}}(u) = \frac{1}{N}\sum_{i}^{N}H(q_i - u)(1 - H(s_i - t)) \tag{8}$$

Formula (8) gives the proportion of samples that have high quality scores, but which ultimately fail recognition.

Annex A
(informative)

Example of encoding a biometric sample quality block

A.1 ASN.1 example

An example of ASN.1 value notation following the abstract syntax defined in ISO/IEC 39794-1 is as follows:

```
qualityBlocks {
  {
    algorithmIdBlock {
      organization 257,
      id 56
    },
    scoreOrError score : 50
  }
}
```

A.2 XML example

An XML example following the XSD defined in ISO/IEC 39794-4 is as follows:

```
<fir:qualityBlocks>
  <cmn:qualityBlock>
    <cmn:algorithmIdBlock>
      <cmn:organization>257</cmn:organization>
      <cmn:id>56</cmn:id>
    </cmn:algorithmIdBlock>
    <cmn:scoreOrError>
      <cmn:score>50</cmn:score>
    </cmn:scoreOrError>
  </cmn:qualityBlock>
</fir:qualityBlocks>
```

A.3 Binary example

A hexdump of quality blocks in tagged binary encoding of finger image quality data block defined in ISO/IEC 39794-4 is as follows:

```
A6 10
  30 0E
    A0 07
      80 02 0101
      81 01 38
    A1 03
      80 01 32
```

Annex B
(informative)

Example of standardized exchange of quality assessment algorithm results

B.1 General

As described in Clause 8, quality assessment algorithm vendors should be able to offer quality measures including raw quality scores and native quality measures (which are not necessarily integers in the range from 0 to 100) to the biometric community. In particular, exchanging quality scores generated from public datasets will be useful in providing technical insight and will allow the consumers of the quality scores to examine and understand how the quality scores relate to the intrinsic information content of samples.

This annex provides an example of exchanging such information in XML format.

B.2 Example quality exchange document

This clause shows an example of an XML coding for vendor “SampleVendor” with id = 123 publishing the results of algorithm “SampleAlgo_v10” with id = 456 on test sets “FERET-grayscale” and “FERET-color”. This example XML document is available at https://standards.iso.org/iso-iec/29794/-1/ed-3.

```
<?xml version="1.0" encoding="UTF-8"?>

<iso:isoVendorQualityReport xmlns:iso="https://standards.iso.org/iso-iec/29794/-1"
  qualityVendorId="123"
  qualityAlgorithmId="456"
  qualityAlgorithmMinValue="0.0"
  qualityAlgorithmMaxValue="100.0">

  <iso:testSets>


    <iso:testSet name="FERET-color" location="https://www.nist.gov/itl/products-and-services/
color-feret-database">
      <iso:sample
      name="ID-00002_931230_fa" qualityValue="51.26"/>
      <iso:sample
      name="ID-00002_931230_fb" qualityValue="82.17"/>
    </iso:testSet>

  </iso:testSets>

</iso:isoVendorQualityReport>
```

B.3 Informative schema for sample XML quality exchange document

This XSD is available at https://standards.iso.org/iso-iec/29794/-1/ed-3.

```
<?xml version="1.0" encoding="utf-8" ?>
<!--Permission is hereby granted, free of charge in perpetuity, to any person obtaining a
copy of the Schema, to use, copy, modify, merge and distribute free of charge, copies of the
Schema for the purposes of developing, implementing, installing and using software based on
the Schema, and to permit persons to whom the Schema is furnished to do so, subject to the
following conditions:

THE SCHEMA IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING
BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND
NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM,
DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM,
```

```
OUT OF OR IN CONNECTION WITH THE SCHEMA OR THE USE OR OTHER DEALINGS IN THE SCHEMA.

In addition, any modified copy of the Schema shall include the following notice:

THIS SCHEMA HAS BEEN MODIFIED FROM THE SCHEMA DEFINED IN ISO/IEC 29794-1, AND SHOULD NOT BE
INTERPRETED AS COMPLYING WITH THAT STANDARD-->
<xs:schema
    xmlns:xs="https://www.w3.org/2001/XMLSchema"
    xmlns:vc="https://www.w3.org/2007/XMLSchema-versioning"
    xmlns="https://standards.iso.org/iso-iec/29794/-1"
    vc:minVersion="1.0"
    targetNamespace="https://standards.iso.org/iso-iec/29794/-1"
    elementFormDefault="qualified"
    attributeFormDefault="unqualified">

  <xs:annotation>
    <xs:documentation xml:lang="en">
      ISO-IEC 29794-1:2024 Vendor Quality Report
    </xs:documentation>
  </xs:annotation>

  <xs:element name="isoVendorQualityReport" type="isoVendorQualityReportType" />

  <xs:complexType name="isoVendorQualityReportType">
    <xs:sequence minOccurs="0">
      <xs:element name="testSets" type="testSetsType"/>
    </xs:sequence>
    <xs:attribute name="qualityVendorId" type="xs:int" use="required"/>
    <xs:attribute name="qualityAlgorithmId" type="xs:int" use="required"/>
    <xs:attribute name="qualityAlgorithmMinValue" type="xs:float" use="required"/>
    <xs:attribute name="qualityAlgorithmMaxValue" type="xs:float" use="required"/>
  </xs:complexType>

  <xs:complexType name="testSetsType">
    <xs:sequence maxOccurs="unbounded">
      <xs:element name="testSet" type="testSetType"/>
    </xs:sequence>
  </xs:complexType>

  <xs:complexType name="testSetType">
    <xs:sequence minOccurs="0" maxOccurs="unbounded">
      <xs:element name="sample" type="sampleType"/>
    </xs:sequence>
    <xs:attribute name="name" type="xs:ID" use="required"/>
    <xs:attribute name="location" type="xs:anyURI" use="required"/>
  </xs:complexType>

  <xs:complexType name="sampleType">
    <xs:attribute name="name" type="xs:ID" use="required"/>
    <xs:attribute name="qualityValue" type="xs:float" use="required"/>
  </xs:complexType>

</xs:schema>
```

Annex C
(informative)

Procedures for aggregation of utility-based quality scores for sample-based systems

C.1 Purpose

This annex 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 samples of 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.

C.2 Method

A hypothetical enterprise collects biometric samples and measures the quality of each using a quality assessment algorithm. The quality scores are expected to be monotonic with mated similarity scores, i.e. higher quality scores result in higher similarity scores. Within this scenario, quality scores are quantized into L levels so that (without loss of generality) q=0, …, L, where q=0 and q=L indicate lowest and highest quality scores, respectively. If the number of biometric samples collected over a given interval in an operational situation is n and this is composed of n_q biometric samples of quality q, then the mean quality across all n samples can be computed. However, arithmetic mean is not the preferred method of summarizing quality scores because all samples, regardless of their quality scores, are given the same weight. If instead the expected utility of a biometric sample t of quality q is $u_q = U(q)$, then a better summary statement of quality is as shown in Formula (C.1):

$$\bar{q} = \frac{\sum_{q=0}^{L} u_q n_q}{\sum_{q=0}^{L} n_q} \tag{C.1}$$

If the utility $u_q(t)$ is actually an estimate of the false match or non-match error rate for samples of quality q of a biometric verification system operating at threshold t, for which error can be estimated, then $\bar{q}\,(t)$ will be an estimate of the expected error rate. Next, a procedure is introduced to compute utility $u_q(t)$ for different levels of a quality assessment algorithm such that the summarized quality score is an estimate of the expected error rate.

Consider a biometric verification system using samples for both probes and references, with a corpus that contains N pairs of biometric samples from N subjects. The first sample represents a reference sample, and the second represents the probe sample.

The samples have integer qualities $q_j^{(1)}$ and $q_j^{(2)}$ for j=1,…, N. Applying V comparison algorithms to the samples provides the following results:

— N mated similarity scores, $s_{jj}^{(v)}$;

— up to $N(N-1)$ non-mated similarity scores, $s_{jk}^{(v)}$ with $j \neq k$

where $v = 1,\ldots,V$ and $V \geq 1$.

a) For each comparison algorithm v and each possible quality score value i from 0 to L, compute $FNMR^v(\tau_v,i)$ in accordance with Formula (C.2). $FNMR^v(\tau_v,i)$is the false non-match rate at decision score threshold τ_v that corresponds to a specific expected system false non-match or false match rate (figure-of-merit or fom, f), of probe samples of quality i with reference samples of quality better than or equal to i using mated scores of comparison algorithm v. Higher quality scores indicate better quality.

for (v = 1,...,V)

for (i = 1,...,L)

$$FNMR^v(\tau_v,i)=\frac{\left|\left\{s_{jj}^{(v)}:s_{jj}\le\tau_v,q_j^{(1)}\ge i,q_j^{(2)}=i\right\}\right|}{\left|\left\{s_{jj}^{(v)}:s_{jj}<\infty,q_j^{(1)}\ge i,q_j^{(2)}=i\right\}\right|} \quad \text{(C.2)}$$

end

end

which results in the following array:

$$\begin{pmatrix} FNMR^1(\tau_1,1)\,FNMR^2(\tau_2,1)\ldots FNMR^v(\tau_v,1) \\ FNMR^1(\tau_1,2)\,FNMR^2(\tau_2,2)\ldots FNMR^v(\tau_v,2) \\ \ldots \\ FNMR^1(\tau_1,L)\,FNMR^2(\tau_2,L)\ldots FNMR^v(\tau_v,L) \end{pmatrix}$$

$V=1$ corresponds to use of a single comparison algorithm. For $V>1$, similarity scores and FNMR should be computed for each comparison algorithm.

b) Compute weight u_i as given in Formula (C.3):

$$u_i(f)=\frac{\sum_{v=1}^{V}FNMR^v(\tau_v,i)}{\sum_{q=0}^{L}\sum_{v=1}^{V}FNMR^v(\tau_v,q)} \quad \text{(C.3)}$$

Thus, the aggregated quality Q across an enterprise is as given in Formula (C.4):

$$Q=\sum_{i=0}^{L}u_i(f)p_i \quad \text{(C.4)}$$

where p_i are fractions of samples with quality q_i. Values of computed Q will not be on a range familiar to users. If all samples were of best quality (i.e. q=L), the result would be Q=u_L. Similarly, the worst case is when all samples in the enterprise are of q=0, which results in Q=u_0. Thus, this formulation would result in quality summaries on the range [u_0, u_L]. Formula (C.4) should be regarded as a general figure of merit for a collection of samples based on overall FNMR across multiple algorithms. However, it is recommended to transform [u_0, u_L] to the recommended range [0,100], which has 0 as the lowest quality and 100 as the best. This can be accomplished by either of the following methods:

1) by relating the quality summary number Q (i.e. expected error rate) back to the native quality range as stated in Formula (C.5):

$$\bar{Q}=U^{-1}(Q)=U^{-1}\left(\sum_{i=0}^{L}u_ip_i\right) \quad \text{(C.5)}$$

where U^{-1} is a function approximation (e.g. piece-wise linear interpolation) of pairs (i,u_i);

2) by mapping (e.g. linear mapping) [u_0, u_L] to [0, 100]. Thus, quality summaries mapped to [0, 100] are given by Formula (C.6):

$$\bar{Q} = 100 \sum_{i=0}^{L} p_i \frac{u_i - u_0}{u_L - u_0} \quad \text{(C.6)}$$

Weights in Formula (C.3) are estimates of the observed false non-match rates computed at some fixed threshold. The result is that these weights are most accurate for that particular threshold and not as accurate for biometric systems operating at other thresholds. In verification applications, where operating threshold is fixed at τ, users of a quality assessment algorithm should follow the outlined procedure to establish dedicated weights.

NOTE Weights in Formula (C.3) are consensus estimates. This means that they were estimated using the observed false non-match rates from a set of comparison algorithms. The result is that the weights are not exactly the weights that would be used for any one algorithm, or for a specified set of algorithms. Weights in Formula (C.3) are regarded as best practice estimates to be used unless other details about the application are known. In verification applications, where a specific set of one or more comparison algorithms are known and available, users of an algorithm are to follow the outlined procedure to establish dedicated weights.

Annex D
(informative)

Example code for computing utility-prediction performance metrics

This Python script for calculating error vs. discard characteristics is available at https://standards.iso.org/iso-iec/29794/-1/ed-3

The lowest tested CPython version is 3.5.6 with numpy version 1.7.0

```
import numpy as np

def compute_edc(pair_comparison_scores, pair_quality_scores, comparison_threshold, comparison_
function):
    """Computes the EDC using `numpy` functions.

    This docstring is written in the numpydoc style.

    Parameters
    ----------
    pair_comparison_scores : array_like
        The pairwise biometric comparison scores (floating-point or integer).
        This does not need to be sorted, but the comparison scores need to correspond to the
`pair_quality_scores`,
        so that `zip(pair_comparison_scores, pair_quality_scores)` represents the pairwise
comparison & quality scores.
    pair_quality_scores : array_like
        The pairwise biometric quality scores (floating-point or integer).
        It is assumed that higher quality scores mean better quality,
        so that comparisons with lower quality scores are discarded first.
        Pairwise quality scores can be derived as the minimum of the pairs' corresponding
sample quality scores.
    comparison_threshold : float or int
        The `pair_quality_scores` are compared against this `comparison_threshold` via the
`comparison_function`
        to determine which of the comparisons represent an "error" in the computed EDC.
    comparison_function : np.ufunc
        Choosing this function depends on whether higher or lower comparison scores mean
higher similarity,
        and it depends on whether a false non-match error vs discard characteristic or a false
match error vs discard characteristic is to be computed.
        The `pair_comparison_scores` are the left-hand side of the comparison,
        `comparison_threshold` is the right-hand side.
        E.g. if higher comparison scores mean higher similarity,
        then `np.less` can be used to compute the false non-match error vs discard
characteristic,
        and `np.greater_equal` can be used to compute the false match error vs discard
characteristic.

    Returns
    -------
    tuple[np.ndarray, np.ndarray]
        The `discard_fractions` and the corresponding `error_fractions` of the computed EDC
are returned.
        For EDC plots, the `discard_fractions` are typically plotted on the x-axis, `error_
fractions` on the y-axis.
    """
    # Create a numpy array and sort it by quality:
    scores = np.array(
        list(zip(pair_comparison_scores, pair_quality_scores)),
        dtype=[("comparison", "f8"), ("quality", "f8")],
    )
    scores.sort(order="quality")
    # Run the EDC computations:
    # The array indices correspond to the discard counts, so 0 comparisons are discarded at
```

```
index 0.
    comparison_count = len(scores)
    # Compute the (binary) per-comparison errors by comparing the comparison scores against
the comparison_threshold:
    error_counts = np.zeros(comparison_count, dtype=np.uint32)
    comparison_function(scores["comparison"], comparison_threshold, out=error_counts)
    # Then compute the cumulative error_counts sum:
    # The total error count will be at index 0, which corresponds to 0 discarded comparisons
(or samples).
    # Conversely, at index comparison_count-1 only one comparison isn't discarded and the
error count remains 0 or 1.
    error_counts = np.flipud(np.cumsum(np.flipud(error_counts), out=error_counts))
    # Usually the EDC should model the effect of discarding samples (instead of individual
comparisons) based on
    # a progressively increasing quality threshold. This means that sequences of identical
quality scores have to be
    # skipped at once. In this implementation the discard counts are equivalent to the array
indices, so computing
    # the relevant array indices for the quality sequence starting points also obtains the
corresponding discard counts:
    discard_counts = np.where(scores["quality"][:-1] != scores["quality"][1:])[0] + 1
    discard_counts = np.concatenate(([0], discard_counts))
    # Subtracting the discard_counts from the total comparison_count results in the remaining_
counts:
    remaining_counts = comparison_count - discard_counts
    # Divide the relevant error_counts by the remaining_counts to compute the error_fractions:
    error_fractions = error_counts[discard_counts] / remaining_counts
    # Divide the discard_counts by the total comparison_count to compute the discard_
fractions:
    discard_fractions = discard_counts / comparison_count
    # Return the discard_fractions together with the corresponding error_fractions:
    return discard_fractions, error_fractions
```

Bibliography

[1] ISO/IEC 19784-1, *Information technology — Biometric application programming interface — Part 1: BioAPI specification*

[2] ISO/IEC 39794-5:2019, *Information technology — Extensible biometric data interchange formats — Part 5: Face image data*

[3] ISO/IEC 19785-1:2020, *Information technology — Common Biometric Exchange Formats Framework — Part 1: Data element specification*

[4] Tabassi E., Olsen M., Bausinger O., Busch C., Figlarz A., Fiumara G., Henniger O., Merkle J., Ruhland T., Schiel C., Schwaiger M., NFIQ 2.0 – NIST fingerprint image quality. NIST Interagency Report 8382, NIST, 2021.

[5] Cappelli R., Ferrara M., Maltoni D., Minutia cylinder-code: A new representation and matching technique for fingerprint recognition. IEEE TPAMI, 32(12), 2010.

[6] Tang Y., Gao F., Feng J., Liu Y., FingerNet: A unified deep network for fingerprint minutiae extraction. In Proc. of IJCB, 2017.

[7] Maio D., Maltoni D., Cappelli R., Wayman J.L., Jain A.K., FVC2000: Fingerprint verification competition. IEEE TPAMI, 24(3), 2002.

[8] Grother P., Tabassi E., Performance of biometric quality measures. IEEE Trans. on Pattern Analysis and Machine Intelligence, **29**(4):531–543, Apr. 2007.

[9] Osorio-Roig D., Schlett T., Rathgeb C., Tapia J., Busch C., Exploring quality scores for workload reduction in biometric identification, in Proceedings of 10th International Workshop on Biometrics and Forensics (IWBF), 2022.

[10] ISO/IEC 8825-1, *Information technology — ASN.1 encoding rules — Part 1: Specification of Basic Encoding Rules (BER), Canonical Encoding Rules (CER) and Distinguished Encoding Rules (DER)*

ICS 35.240.15
Price based on 26 pages

iso.org