International Standard

ISO/IEC 29794-5

Information technology — Biometric sample quality —

Part 5: Face image data

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

Partie 5: Données d'image de face

First edition
2025-04

Reference number
ISO/IEC 29794-5:2025(en)

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 first edition cancels and replaces the first edition of ISO/IEC TR 29794-5:2010 which has been technically revised.

The main changes are as follows:

— the document has been completely revised to become and International Standard;

— information on the role of quality measures has been added;

— requirements on quality software have been added.

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

Adoption of deep learning techniques has caused error rates associated with automated face recognition tasks to be reduced. However, errors still occur and are often related to imaging, human factors, the level of biometric capture subject cooperation, the comparison algorithm, and its associated threshold and decision logic. Without significant modernisation of capture procedures, recognition errors will become more prevalent as volumes increase. This document is aimed at reducing errors due to image quality, through the use of quality assessment algorithms. Quality assessment algorithms have several roles (see Annex C), primarily those related to sample capture. Drivers for improved capture are as follows.

— Need for improved usability — The general improvement of biometric systems has highlighted that improved usability for both biometric capture subjects and human operators can reduce errors through the improvement of capture. Without a careful consideration of both biometric capture subjects and system operators, system designers risk seeing the limitations inherent in using technology alone.

— Increasing volumes — Vast numbers of face images are being collected in many commercial, civil identity management and law enforcement applications. These photographs are used as reference enrolment samples, or as recognition probes that, in turn, sometimes later serve as references.

— New programs — Future large-scale programs will employ face recognition: For example, in China the railway transportation system uses face recognition for identity verification and to improve passenger check-in efficiency. The European Union uses face recognition for biometric exit confirmation. The United States currently uses face recognition for biometric exit confirmation and vessel boarding. In India, the Aadhaar program allows face recognition for authentication.

— Face-blind cameras — Historically, many face images were collected using cameras that were not face-aware. In contrast, in some situations concerning fingerprint and iris biometrics, capture devices run in an auto-capture quality-assessment loop, with explicit awareness of the kind of image intended for collection.

— Reliance on imaging design specifications — Faces collected for ID credentials and authoritative databases are largely collected using cameras set up according to published documentary standards, most recently ISO/IEC 39794-5, regulating geometry and photography. In the best case, face images from such collections are then checked with image compliance tools. When photographs are collected by a human photographer, this can be without any automated quality assessment, relying only on the photographer to check conformance.

— Behaviour not intended by the relevant capture standard — Some recognition failures arise from biometric capture subjects effecting differences in presentation in reference and probe images. Standards define a canonical presentation to be centred and frontal with neutral expression, eyes-open and without occlusions. Facial recognition systems are expected to operate accurately across a wide range of individuals who vary in age, body size, ethnicity, language, culture, literacy and familiarity with technology. Careful human factors design is vital to the acquisition of canonical images and improved face image capture.

— Quality assessment is separated from the capture process — In many cases, a photograph is captured and later submitted to a backend server while ensuring no image tampering occurs, where it is assessed for quality. If poor quality is detected (by human or automated means), re-capture is initiated hours or days later, when possible, with another encounter and attendant expense.

Regarding image quality, Table 1 lists characteristics of face image quality relating to the biometric capture subject and characteristics relating to the capture process, demonstrating that issues due to mis-presentation (often associated with human factors design) and issues related to imaging are in many cases separable. For example, photographs can be systematically de-focused even when the biometric capture subjects present perfectly.

Table 1 — Characterization of face image quality

	Biometric capture subject characteristics	Capture process
Static properties	Biological characteristics; — injuries and scars, — dermatological conditions, — etc.	Capture process and capture device properties: — image resolution, — optical distortions, — sub-optimal camera angle, — field of view, — etc.
	Other static characteristics: — thick or dark glasses, — permanent jewellery, — makeup and cosmetics, — etc.	Static properties of the background: — (textured) wallpaper.
		Affordance: — properties of a data capture subsystem that intuitively imply its functionality and use to biometric capture subjects, — human-centric system physical and process design.
Dynamic properties	Behaviour: — exaggerated expression, — hair across the eye, — facial hair, — etc.	Scenery: — background moving objects, — variation in lightning.
		Capture device variation: — de-focus, — camera vibration, — sub-optimal camera angle, — poor exposure, — etc.

By defining image quality measurements, this document is intended to improve the accuracy of automated face recognition systems. Quality can be tied to recognition accuracy (see Annex B). Improved quality can also improve human review of images. The quality measures included in this document were selected because guidance on how to control them has already been included in ISO/IEC 39794-5. The implementations of some quality measures were evaluated for performance.[62] The reference implementation defines quality measures that use external algorithms with licence conditions.[58]

This document recognizes the Open Face Image Quality (OFIQ)[60] software as the reference implementation of the requirements of the document. It is open-source.[59] Other quality algorithm implementations can conform to this document as described in Clause 5.

Some of the computations of this document can be effective on images captured with illumination at non-visible wavelengths.

Encoding of quality data is defined in ISO/IEC 29794-1. The methodology for performance evaluation of quality assessment algorithms is also defined in ISO/IEC 29794-1.

NOTE Use of this document can be subject to local regulations.

Information technology — Biometric sample quality —

Part 5:
Face image data

1 Scope

This document establishes requirements on implementations that quantify how a face image's properties conform with those of canonical face images, for example those specified in ISO/IEC 39794-5:2019, Clause D.1, for three use-cases:

1) collection of reference samples for ID documents;

2) sample system enrolment; and

3) probes for instantaneous response.

This document also establishes terms and definitions for quantifying face image quality and specifies methods for quantifying the quality of face images.

This document does not establish requirements on:

— assessing the quality of pairs or sequences of images;

NOTE This document establishes requirements for software that inspects exactly one image. This document does not establish requirements for software that compares two or more images (such as biometric recognition). However, the computations of this document can be applied separately to each image in a pair or sequence.

— assessing the quality of 3D captures;

— encodings of face image quality data;

— performance evaluation of face image quality assessment algorithms.

The use cases within scope of this document primarily address the assessment of images from data capture subjects who consent to processing of their biometric data, or for whom biometric capture is operationally authorized.

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 29794-1, *Information technology — Biometric sample quality — Part 1: Framework*

ISO/IEC 39794-1:2019, *Information technology — Extensible biometric data interchange formats — Part 1: Framework*

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

ISO/IEC 19794-1:2011, *Information technology — Biometric data interchange formats — Part 1: Framework*

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

IEC 61966-2-1:1999, *Multimedia systems and equipment — Colour measurement and management — Part 2-1: Colour management — Default RGB colour space — sRGB*

IEC 61966-2-2:2003, *Multimedia systems and equipment — Colour measurement and management — Part 2-2: Colour management — Extended RGB colour space — scRGB*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 2382-37, ISO/IEC 29794-1, ISO/IEC 39794-5 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
canonical face image
face image conformant to an external standard or specification of a reference face image

Note 1 to entry: In many applications, the canonical face image is that given in ISO/IEC 39794-5:2019, Clause D.1, which specifies a reference face image for a machine-readable travel document.

Note 2 to entry: Most of the computations of this document can be effective on images captured in automated border control gates, visa images and other server-side images, which are in scope of this document.

3.2
de-focus
aberration in which an image or part of an image is out of focus

Note 1 to entry: De-focus tends to reduce the sharpness and contrast of the image.

3.3
face detection
process of determining whether and where faces are present in an image

3.4
face image
electronic image-based representation of the face of a capture subject

Note 1 to entry: Any image captured for use-cases 1 – 3 described in Clause 5 is considered as a face image.

Note 2 to entry: ISO/IEC 39794-5 includes a definition for face portrait as the visual representation of the capture subject, which includes the full-frontal part of the head, including hair in most cases, as well as neck and possibly top of shoulders. Face portraits appear in several places on and in a machine-readable travel document (MRTD).

Note 3 to entry: Given an image that has a roll angle of 90 ° or more (which is far from the presentation intended by ISO/IEC 39794-5:2019), a QAA can assign low quality component values or fail to return a record

[SOURCE: ISO/IEC 39794-5:2019, 3.27, modified — Notes to entry have been added.]

3.5
face bounding box
rectangle containing the central region of interest of a face visible in the face image

Note 1 to entry: The face bounding box is used for the estimation of landmarks to restrict the face image to the region of interest.

Note 2 to entry: The face bounding box is the result of the face detection process as defined in 6.4.

3.6
face landmarks
set of anthropometric points in the image marking the contour and different parts of the face

Note 1 to entry: The face landmarks are computed by the algorithm in 6.5.

3.7
landmarked region
minimal convex region of the face image containing all face landmarks

Note 1 to entry: The landmarked region of the face is defined by the landmark estimation algorithm, as described in 6.5. It encompasses the area from (and including) eyebrows to chin and from (but excluding) left ear to right ear.

Note 2 to entry: The inner region, as defined in ISO/IEC 39794-5:2019, 3.39 and D.2.2, is a coarse approximation of the landmarked region.

3.8
face alignment
process of rotating, translating and scaling a face image so that the transformed image has certain dimensions and the eyes, nose and mouth corners are approximately located at pre-defined locations in the transformed image, and applying the same transformation to the face landmarks

3.9
face parsing
process of assigning semantic labels to regions contained in a face image specifying the part of the subject depicted in the region

Note 1 to entry: The set of parts to which the pixels are assigned comprises various parts of the face, neck, hair, clothing and various accessories. Pixels not belonging to the subject (background) are assigned to the value 0.

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

[SOURCE: ISO/IEC 29794-1:2024, 3.10]

3.11
pose estimation
process of determining the 3-axis rotation of the head in an image

Note 1 to entry: Pose estimation requires a specified coordinate system for definition of the angles and their sense (left hand vs. right hand). The default choice is given in ISO/IEC 39794-5:2019, 7.21.

3.12
sharpness
clarity of fine details in an image

Note 1 to entry: Sharpness will be improved with precise focus, good imaging resolution, and absence of biometric capture subject motion relative to the camera.

3.13
ICC profile
International Color Consortium profile
set of data that characterizes colour input or output by defining the mapping between the data and a profile defined colour space

Note 1 to entry: ICC profiles (e.g. sRGB and ROMM RGB) are intended to provide a standard approach to colour management needs.

Note 2 to entry: The captured face portrait is a true-colour representation of the holder in a typical colour space such as sRGB. Other true-colour representations, such as Adobe RGB (1998) or ProPhoto RGB (ROMM RGB), are used as specified in ISO/IEC 39794-5: 2019, D.1.4.2.9.

3.14
quality assessment algorithm
quality algorithm
algorithm to calculate a quality measure

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

[SOURCE: ISO/IEC 29794-1:2024, 3.13]

4 Abbreviated terms

For the purposes of this document, the acronyms and abbreviated terms given in Table 2 apply.

Table 2 — Acronyms and abbreviated terms

Acronym/ abbreviated term	Definition
BGR	colour space with the order of colours (blue, green, red)
BU	background uniformity
CNN	convolutional neural network
DFT	discrete Fourier transform
EVZ	eye visibility zone
ICC	International Color Consortium
IED	inter-eye distance
IM	illumination mean
IU	illumination uniformity
NaN	not a number
px	pixels
QAA	quality assessment algorithm
QC	quality component
QCV	quality component value
OFIQ	Open Source Face Image Quality
QM	quality measure
OpenCV	OpenCV version 4.5.5
QS	(unified) quality score
RGB	colour space with the order of colours (red, green, blue)
UC	use-case

5 Conformance

A claim of conformance to this document is made by asserting support for one or more use-case(s) (see below) and a dated edition of this document.

To conform with this document, a face image quality assessment implementation shall:

— implement the computations marked as mandatory (M) in Table 3 for the claimed use case(s);

— conform to the requirements of Clauses 6 and 7;

— conform to the requirements of Clause 8 (face image quality block encoding); and

— meet the requirements specified in Annex A.

Conformance to the requirements of Clause 8 fulfils Level 1 and Level 2 conformance as specified in ISO/IEC 39794-1:2019, Annex C. Conformance to the requirements of 7.2, 7.3 and 7.4 fulfils Level 3 conformance as specified in ISO/IEC 39794-1:2019, Annex C.

NOTE 1 Organizations deploying quality assessment algorithms in the various use cases can choose to compute the provided quality measures and to relate their values to any use-case specific thresholds.

NOTE 2 For standardized interchange of quality values produced by implementations of this document, ISO/IEC 29794-1 defines a standardized quality block and requires quality measures to be an integer between 0 and 100.

Three use-cases are considered:

— UC1: Collection of reference samples for ID documents. The face image will be stored on a document, used for example for a maximum of 10 years, and should support human examination.

— UC2: System enrolment, current or later creation of a reference, delayed recognition. Acquisition of face images where quality should be high enough to ensure later usage and interoperability.

— UC3: Collection of probe samples for instantaneous recognition. Single use of a face image with instantaneous response.

Table 3 — Conformance requirements by activity

#	Face image quality measure	Sub-clause	UC1 (ID Documents)	UC2 (System enrolment)	UC3 (Probe for instantaneous recognition)
1.	Quality score (unified)	7.2	M	M	M
2.	Background uniformity	7.3.2	M	O	O
3.	Illumination uniformity	7.3.3	M	O	O
4.	Luminance mean	7.3.4.2	O	O	O
5.	Luminance variance	7.3.4.3	M	O	O
6.	Under-exposure prevention	7.3.5	O	O	O
7.	Over-exposure prevention	7.3.6	O	O	O
8.	Dynamic range	7.3.7	M	O	O
9.	Sharpness	7.3.8	M	O	O
10.	No compression artefacts	7.3.9	O	O	O
11.	Natural colour	7.3.10	O	O	O
12.	Single face present	7.4.2	M	M	O
13.	Eyes open	7.4.3	M	O	O
14.	Mouth closed	7.4.4	M	M	O
15.	Eyes visible	7.4.5	M	M	O
16.	Mouth occlusion prevention	7.4.6	M	M	O
17.	Face occlusion prevention	7.4.7	M	M	O
18.	Inter-eye distance	7.4.8	M	M	M
19.	Head size	7.4.9	M	M	M
20.	Leftward crop of face in image	7.4.10.1	M	M	M
21.	Rightward crop of face in image	7.4.10.2	M	M	M
22.	Margin above face in image	7.4.10.3	M	M	M
23.	Margin below face in image	7.4.10.4	M	M	M

Key

M mandatory for the QAA to implement this quality measure

O optional

Table 3 *(continued)*

#	Face image quality measure	Sub-clause	UC1 (ID Documents)	UC2 (System enrolment)	UC3 (Probe for instantaneous recognition)
24.	Pose angle yaw frontal alignment	7.4.11.2	M	M	O
25.	Pose angle pitch frontal alignment	7.4.11.3	M	M	O
26.	Pose angle roll frontal alignment	7.4.11.4	M	M	O
27.	Expression neutrality	7.4.12	M	O	O
28.	No head covering	7.4.13	M	O	O
29.	Radial distortion	D.2.1	O	O	O
30.	Pixel aspect ratio	D.2.2	O	O	O
31.	Gaze	D.3.1	O	O	O
32.	Shoulder presentation	D.3.2	O	O	O
33	Camera subject distance	D.3.3	O	O	O
34	Motion blur prevention	D.3.4	O	O	O
Key					
M mandatory for the QAA to implement this quality measure					
O optional					

6 Common computations

6.1 Overview

Quality measures defined in this document can be applied to either 8 bit encoded greyscale images, or 24 bit encoded colour images. Clause 6 supports the computations in Clause 7 by defining certain steps that appear in more than one of its subclauses.

Several subclauses of Clauses 6 and 7 reference the usage of the pre-trained neural networks listed in Table 4. For each, their implementation description, model, weights, and alternative implementations, are provided in Reference [58].

Table 4 — List of neural networks used in computations of Clauses 6 and 7

#	Clause	Model	Implementation	Dataset
1	6.4	Face detection	See Reference [30]	NA
2	6.5	Face landmark estimation	See Reference [53]	Wider Facial Landmarks in the Wild (WFLW) See Reference [31]
3	6.8	Face parsing	See Reference [68]	CelebMask-HQ[69]
4	6.9	Face extraction	See Reference [32]	CelebMask-HQ[69]
5	6.11	Head pose estimation	See Reference [50]	300W-LP[70]
6	7.2	Unified quality scoring	See Reference [54], [55]	MS1MV2[54]
7	7.3.9.2	No Compression artefacts	See Reference [58]	OFIQ Development Dataset[60]
8	7.4.12.2	Expression neutrality	See References [9], [51], [52]	OFIQ Development Dataset[9], [51], [52], [60]

6.2 Conversion of 16 bits per channel images to 8 bits per channel images

When 16-bit scRGB images are given, conversion of 16-bit scRGB images shall be performed as defined in IEC 61966 2-1, using conversions as defined in IEC 61966-2-2:2003, Clause A.2.

NOTE IEC 61966-2-2:2003, 3.1 and 4.1 show the conversion chain for 16-bit scRGB to 1931 CIEXYZ to 8-bit sRGB.

Images with other ICC profiles shall be converted according to their profile.

6.3 Conversion of high bit-depth images to 8 bit greyscale or 24 bit colour images

ISO/IEC 39794-5:2019 allows high-bit depth face captures with native camera formats that are encoded as defined in ISO/IEC 39794-5:2019, D.1.4.2.4. ISO/IEC 39794-5:2019 allows also proprietary formats with non-normative encoding specifications. Conversion, such as scale adaptation, should follow manufacturer recommendation.

Conversion of high bit-depth images to 8 bits greyscale or 24 bits colour images shall be completed before quality measures are computed. In the conversion, it is recommended to follow the methods for conversion of high dynamic range content to standard dynamic range described in Reference [47].

6.4 Face detection

This subclause supports the computations of Clause 6 and Clause 7 that depend on detection of faces in a face image. Face detection can be performed using a neural network pre-trained to produce face bounding boxes using face image datasets annotated with face bounding boxes. The face detection algorithm should accept any image and return a set of face bounding boxes for each face detected in the image.

The implementation should include the algorithm below. It may include an alternative detection algorithm if the resulting Clause 7 quality outputs are reproduced according to Annex A. The algorithm takes as input a face image I in the BGR colour channel order (8 bits per channel). It uses the model from Reference [30]. The model is available from Reference [58].

1) Determine the height, h, and the width, w, of image I.

2) Pad image I from the left and right side by 0.2 w and from the top and bottom by 0.2 h, and update w and h to the new width and height of I.

3) Resize image I to size (height, width) 300 px × 300 px using OpenCV with bilinear interpolation.

4) Normalize the values of I using mean μ = (104.0, 117.0, 123.0) and standard deviation σ = (1, 1, 1) using 6.19 to obtain the array A.

5) Encode A as input tensor $T^{[0]}$ of dimensions (1, 3, 300, 300) for the neural network model.

6) Run a forward pass through the model using $T^{[0]}$ as input to obtain the output tensor $T^{[1]}$ of dimensions (1, 1, N, 7), where N is the number of detected faces prior to filtering.

7) Extract the last 2 dimensions of $T^{[1]}$ as face bounding box table F of size (N, 7), where each row of F contains the data of the detected face bounding boxes:

    a) column 3 specifies the confidence;

    b) columns 4 to 5 specify the relative (i.e. scaled to the interval [0,1]) x- and y-coordinates, respectively, of the upper left corner;

    c) columns 6 to 7 specify the relative (i.e. scaled to the interval [0,1]) x- and y-coordinates, respectively, of the lower right corner.

8) Delete all rows in F, where one of the following conditions is fulfilled:

    a) the confidence is lower than 0.4, i.e. $F_{i,3} \leq 0.4$;

b) the width of the face bounding box is smaller than 1/20 of the image width, i.e. $F_{i,6}$ - $F_{i,4}$ < 0.05;

c) the face bounding box protrudes outside the image boundary, i.e. $F_{i,4} \leq 0$ or $F_{i,6} \geq 1$ or $F_{i,5} \leq 0$ or $F_{i,7} \geq 1$.

9) If no rows remain, the abstract value failureToAssess is returned.

10) For each row i, obtain the coordinates (a_i,b_i) and (c_i,d_i) of the upper left corner and the lower right corner of the face bounding box within the input image, respectively, as:

$$a_i = \lfloor w \cdot F_{i,4} - p_1 + 0.5 \rfloor, \quad b_j = \lfloor h \cdot F_{i,5} - p_2 + 0.5 \rfloor,$$

$$c_i = \lfloor w \cdot F_{i,6} - p_1 + 0.5 \rfloor, \quad d_j = \lfloor h \cdot F_{i,7} - p_2 + 0.5 \rfloor.$$

11) Determine the row j with the largest face bounding box area, i.e.:

$j = \mathrm{argmax}((c_j - a_j)\cdot(d_j - b_j))$.

If there are several face bounding boxes with largest area, choose the first one of these.

12. Return the face bounding box coordinates (a_j, b_j, c_j, d_j).

As an example, Figure 1 shows the typical output for a face image with significant roll angle. The image is selected specifically to not be a canonical face image.

NOTE 1 The coordinates (a_i, b_i, c_i, d_i) computed in step 11 can lie outside the image region, i.e. $a_i \geq 0, b_i \geq 0, c_i \leq w-1, d_i \leq h-1$ does not necessarily hold. The same holds for the coordinates of the largest face bounding box output by the algorithm.

NOTE 2 The face detection definitions of this subclause accommodate images where one or two eyes are visible, or can be closed or occluded, and images where multiple faces are present.

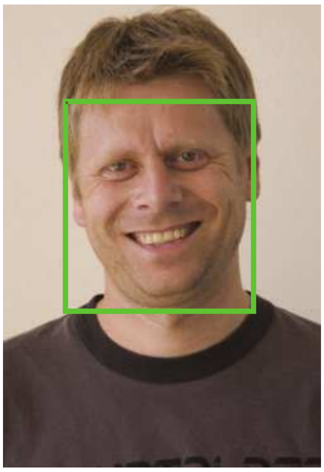


Figure 1 — Example face bounding box

6.5 Face landmark estimation

This subclause supports the computations of Clause 6 and Clause 7 that either require specific landmarks or the landmarked region (see Figure 2 for an illustration of location and naming of landmarks in an example face image). Face landmark estimation can be performed using a neural network pre-trained to compute the position of face landmarks using face image datasets annotated with face landmarks to be estimated.

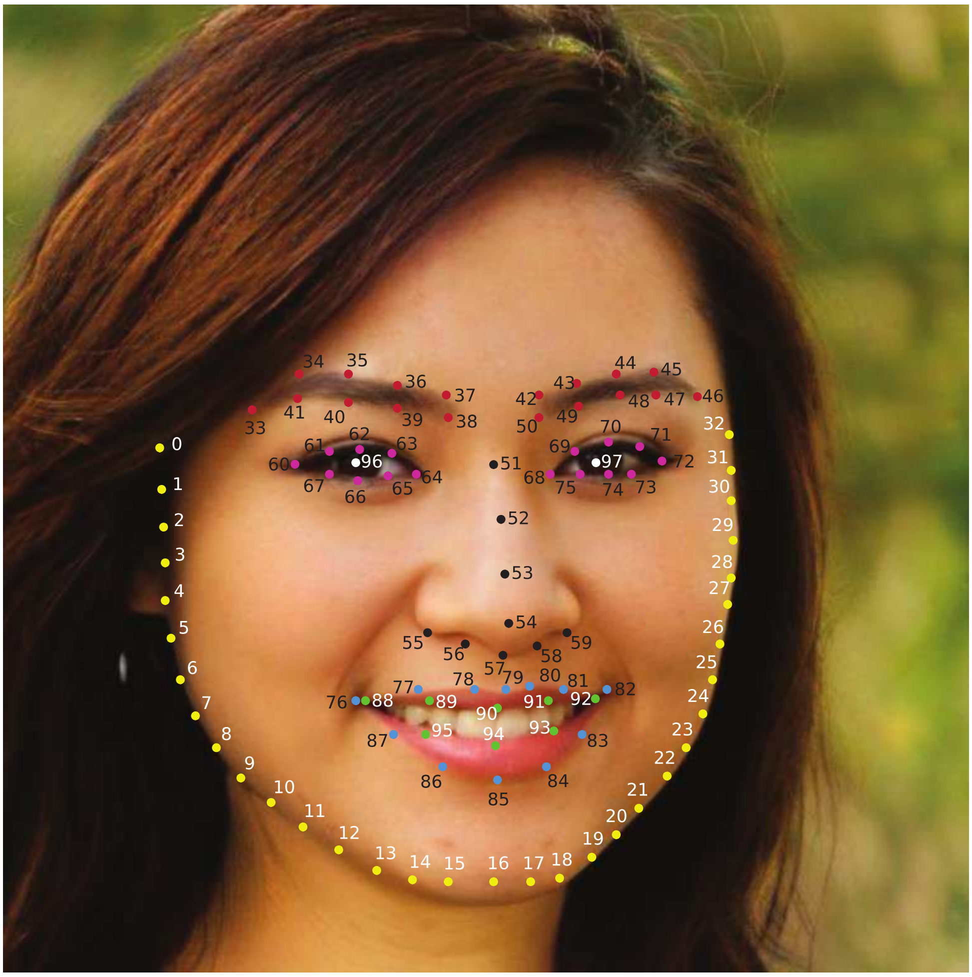


Figure 2 — Location and indices of landmarks in an example face image

The landmark estimation algorithm should accept an image along with a face bounding box returned by face detection. The estimation algorithm should return the 98 face landmarks (see the example in Figure 2).

An implementation should include the algorithm below. It may include an alternative landmark estimation algorithm if the resulting Clause 7 quality outputs are reproduced according to Annex A.

The algorithm uses the ADNET CNN[53] which has been trained on the Wider Facial Landmarks in the Wild (WFLW) dataset.[31] The model is available from Reference [58]. These face landmarks are used instead of those referred to in ISO/IEC 39794-5:2019 (i.e. the anthropometric landmarks of ISO/IEC 14496-2:2004), because there is no robust implementation for computing the latter landmarks.

Table 5 — Meaning of the colours of the landmarks in Figure 2

Colour	Labelled part
Yellow	Face contour
Red	Eye brows
Black	Nose
Purple	Outer boundary of eyes
Blue	Outer boundary of lips
Green	Inner boundary of lips
White	Pupils

The algorithm takes as input a face image I in BGR colour channel order (8 bits per channel) and the face bounding box (a, b, c, d) output by the face detection algorithm specified in 6.4.

1) Determine the height, h, and the width, w, of I.

2) Extend the face bounding box on both sides to square shape:

   a) if $c - a < d - b$, then update $(a,c) \leftarrow \left(\left\lfloor \frac{a+c}{2} - \frac{d-b}{2} \right\rfloor, \left\lfloor \frac{a+c}{2} + \frac{d-b}{2} \right\rfloor \right)$;

   b. if $c - a > d - b$, then update $(b,d) \leftarrow \left(\left\lfloor \frac{b+d}{2} - \frac{c-a}{2} \right\rfloor, \left\lfloor \frac{b+d}{2} + \frac{c-a}{2} \right\rfloor \right)$.

3) Determine the necessary padding as:

   a) $p_{\mathrm{LEFT}} = \max(0, -a)$,

   b) $p_{\mathrm{RIGHT}} = \max(0, c - w + 1)$,

   c) $p_{\mathrm{TOP}} = \max(0, -b)$,

   d) $p_{\mathrm{BOTTOM}} = \max(0, d - h + 1)$.

4) Pad the image with black colour (0, 0, 0) from left, top, right, bottom side by p_{LEFT}, p_{TOP}, p_{RIGHT}, p_{BOTTOM} pixels, respectively.

5) Update a, b, c, d to $a + p_{\mathrm{LEFT}}$, $b + p_{\mathrm{TOP}}$, $c + p_{\mathrm{LEFT}}$, $d + p_{\mathrm{TOP}}$.

6) Crop I to (a, b, c, d), i.e. to $I[a{:}c, b{:}d]$.

7) Resize I to size (256, 256).

8) Normalize the values of I using mean $\mu = 127.5$ and standard deviation $\sigma = 127.5$ as specified in 6.19 to obtain the array A.

9) Encode A as input tensor $T^{[0]}$ of dimensions (1, 256, 256).

10) Run a forward pass through the model using $T^{[0]}$ as input to obtain the output tensor $T^{[1]}$ of dimensions (1, 98, 2).

11) Reshape $T^{[1]}$ to an array A of dimensions (98, 2).

12) De-normalize A to obtain the array B specifying the landmark positions within the cropped and resized image by setting:

    $B_{i,j} = 127.5\,(A_{i,j} + 1)$ for all i, j.

13) Compute the list of landmarks (x_i, y_i) within the original input image I as:

    $x_i = ROUND\left(B_{i,1} \frac{d-b}{256} + a - p_{\mathrm{LEFT}} \right)$ and $y_i = ROUND\left(B_{i,2} \frac{d-b}{256} + b - p_{\mathrm{TOP}} \right)$.

14) Output all landmarks, $L_0 = (x_0, y_0), \ldots, L_{97} = (x_{97}, y_{97})$.

6.6 Landmarked region segmentation

The algorithm takes as input a set of landmarks, either the original landmarks computed by the algorithm in 6.5 or the transformed landmarks computed by the algorithm in 6.7.

1) Compute the convex hull of the landmarks.

2) Generate a segmentation map S as grayscale image (one channel), where all pixels inside and on the convex hull are set to 1 and all pixels outside are set to 0.

3) Return S.

An example of the segmented map appears in Figure 3; note the forehead is excluded.

Figure 3 — Example landmarked region

6.7 Face alignment

This subclause supports computations in Clauses 6 and 7 that require the position of the eyes, nose and mouth to be roughly at a pre-defined position. Face alignment is the translation, rotation and scaling of the image to do this, as illustrated in Figure 4.

Figure 4 — Example aligned face

The face alignment algorithm takes as input the image I and corresponding facial landmarks L_0, ..., L_{97} and then performs the following steps:

1) For both eyes, compute the eye centres C_r, C_l as specified in 6.10.

2) Determine the position L_{54} of the nose tip.

3) Determine the positions $M_r = L_{76}$ and $M_l = L_{82}$ of the subject's right and left mouth corner.

4) Determine the isogonal affine transformation that optimizes the mapping of the points C_r, C_l, L_{54}, M_r, and M_l to the target points (251, 272), (364, 272), (308, 336), (262, 402), and (355, 402), respectively. The OpenCV function estimateAffinePartial2D should be used with the "least median square" approach (method=LMEDS).

5) Apply the transformation T to the image I to obtain a transformed image I' then crop and/or pad to dimensions of 616 px × 616 px. For padding, use black colour (0, 0, 0).

6) Apply the transformation T to the landmarks L_0, ..., L_{97} to obtain transformed landmarks L'_0, ..., L'_{97}.

7) Output the transformed image I', the transformed landmarks L'_0, ..., L'_{97} and the transformation matrix T.

6.8 Face parsing

Face parsing supports some computations in Clause 7 by generating a segmentation map. Face parsing can be performed using a neural network pre-trained to label pixels in a face image that correspond to different face parts or classes using face image datasets annotated with such labels. The face parsing network should accept an aligned face image and return a segmentation map corresponding to the 19 classes defined in Table 5. An example map is shown in Figure 5.

Figure 5 — Example segmentation map

Table 6 — Label values with their body part or region

Index	Feature	Index	Feature
0	background	10	nose
1	face skin	11	mouth
2	left eye brow	12	upper lip
3	right eye brow	13	lower lip
4	left eye	14	neck
5	right eye	15	necklace
6	eyeglasses	16	clothing
7	left ear	17	hair
8	right ear	18	head covering
9	earring		

Some body parts and regions listed in Table 6 are illustrated in the example of Figure 5.

The implementation should include the algorithm below. It may include an alternative face parsing algorithm if the resulting Clause 7 quality outputs are reproduced according to Annex A.

The algorithm takes as input an image I in RGB colour channel order (8 bits per channel) that has been aligned as specified in Clause 7 and uses the model[32][53] converted from the original PyTorch model available from Reference [58].

1) Crop image I by 30 pixels from both sides and by 60 pixels from the bottom.

2) Scale image I to 400 px × 400 px using bilinear interpolation to obtain image I'.

3) Normalize I' with mean (123.7, 116.3, 103.5) and standard deviation (58.4, 57.1, 57.4) as specified in 6.19.

4) Reshape I' in a 4-dimensional tensor of size (1, 3, 400, 400).

5) Run a forward pass through the face parsing model to obtain an output tensor of size (19, 400, 400).

6) Compute the argmax on first axis to obtain the segmentation map S of size 400 × 400.

7) Return S.

NOTE For UC1, ISO/IEC 39794-5:2019, D.1.5.4, requires that only down-sampling be allowed as post-processing of captured images, which are stored in face portraits in identity documents. The upscaling by bilinear interpolation, as an element of this algorithm, is performed for quality assessment.

6.9 Face occlusion segmentation

Face occlusion segmentation supports computations in Clause 7 that measure the occlusion of the face or visibility of mouth and eyes. It computes a segmentation map marking the parts of the face that are not occluded (e.g. by hair) as illustrated in purple regions of Figure 6.

Figure 6 — Visualization of the segmentation (purple colour) of the regional occluded face image given by the example image c-04-headcovered

The implementation should include the algorithm below. It may include an alternative face occlusion detection algorithm if the resulting Clause 7 quality outputs are reproduced according to Annex A.

Face occlusion segmentation can be performed using a neural network pre-trained to produce a segmentation map that labels all image pixels corresponding to the face sample which are not occluded by other objects (e.g. by hats, glasses or hair). The face occlusion segmentation network is trained using face image datasets annotated with such labels.

The occlusion segmentation algorithm uses the following definitions of what is an occlusion:

1) Non-beard, non-moustache, non-eyebrow, hair that occludes facial skin.

2) Hair on the forehead hanging over the eyes.

3) A tongue poked out of the mouth.

4) Makeup if there is a visible border to face skin, which is only the case if the makeup has a clearly unnatural colour.

5) Opaque lenses of spectacles.

6) Frames of eyeglasses.

7) Specular reflections from eyeglasses.

8) Any objects (for example, hands or protective faces masks) are considered as occlusion.

The following are not considered occlusions:

1) Transparent lenses of spectacles.

2) Beards and moustaches.

3) Eyebrows.

4) At extreme pose angles, the face can self-occlude but this is not considered occlusion.

The algorithm takes as input an aligned image I in RGB colour channel order (8 bits per channel) that has been aligned as specified in 6.7 and uses a CNN model[58] converted from the original PyTorch model,[32] which itself is based on References [33], [34] and [35]:

1) Crop the image from all sides by 96 pixels.

2) Scale the image to 224 px × 224 px using bilinear interpolation to obtain image I.

3) Encode the image in a 4-dimensional tensor of size (1, 3, 224, 224).

4) Divide the tensor by 255.

5) Feed the tensor through the face occlusion segmentation CNN to obtain an output tensor of size (1, 224, 224).

6) Remove the first dimension of the output tensor to obtain a segmentation map M of size (224, 224).

7) Set all positive values of the segmentation map M to 1, and all other values to 0.

8) Resize the segmentation map M to 424 px × 424 px using nearest neighbour interpolation (using truncation for rounding non-integer indices).

9) Pad the segmentation map M by 96 pixels from all sides setting the padded regions to 0 indicating absence of face or occluded face.

10) Output the segmentation map M.

6.10 Computing eye centres and inter-eye distance

Inter-eye distance is used to give an estimate of the resolution of a face image.

The algorithm takes as input the landmarks $L_0, \ldots, L_{97}$ estimated by the algorithm in 6.5 and the head pose angle for the yaw, ϕ_{yaw}, computed by the algorithm in 6.11.

1) The coordinates of the centre of the subject's left eye, $(X_{\text{L}}, Y_{\text{L}})$, are computed as the arithmetic mean of the coordinates of landmarks L_{68} and L_{72} marking the inner and outer canthi of the left eye, respectively, i.e. $(X_{\text{L}}, Y_{\text{L}}) = (L_{68} + L_{72})/2$.

2) The coordinates of the centre of the subject's right eye, $(X_{\text{R}}, Y_{\text{R}})$, are computed as the arithmetic mean of the coordinates of landmarks L_{60} and L_{64} marking the inner and outer canthi of the right eye, respectively, i.e. $(X_{\text{R}}, Y_{\text{R}}) = (L_{60} + L_{64})/2$.

3) Compute d_{IED}, the yaw-corrected Euclidean distance between the left and right eye centres:

$$d_{\text{IED}} = \begin{cases} NaN, & \textit{if } \cos\phi_{\text{yaw}} < \varepsilon \\ \dfrac{1}{\cos\phi_{\text{yaw}}} \sqrt{(X_{\text{L}} - X_{\text{R}})^2 + (Y_{\text{L}} - Y_{\text{R}})^2}, & \textit{if } \cos\phi_{\text{yaw}} \geq \varepsilon \end{cases},$$

where the tolerance value $\varepsilon = 10^{-6}$.

3. Output $(X_{\text{L}}, Y_{\text{L}})$, $(X_{\text{R}}, Y_{\text{R}})$ and d_{IED}.

6.11 Head pose estimation

An implementation shall be capable of estimating three head orientation angles (yaw, pitch and roll) from a face image. An implementation should include the algorithm below. It may include an alternative face pose estimation algorithm if the resulting Clause 7 quality outputs are reproduced according to Annex A.

Face pose estimation can be performed using a neural network pre-trained to estimate head orientation (yaw, pitch, roll) using face image datasets annotated with ground-truth pose values. The face pose estimation network should accept an image and a face bounding box and return estimated yaw, pitch, and roll values.

The implementation should use the following algorithm which employs a pose estimator,[49] the model for which is available at Reference [58] and is a copy of the original implementation.[50] The model estimates three head orientation angles (yaw, pitch and roll) from a face image.

The algorithm takes as input an image I in BGR colour channel order (8 bits per channel) and the coordinates (a,b,c,d) of the largest face bounding box and performs the following steps.

1) Compute the height $l = d - b$ and the centre $(x,y)=\left(\frac{a+c}{2},\frac{b+d}{2}\right)$ of the face bounding box.

2) Set $b=\lfloor y-0.44\cdot l\rfloor$ and $d=\lfloor y+0.51\cdot l\rfloor$.

3) Extend the box to a square by setting $a=\lfloor x+0.5(d-b)\rfloor$ and $c = a + d - b$.

4) If the new face bounding box (a,b,c,d) is not within the image region $[0,w-1]\times[0,h-1]$, where w and h are the width and height of I, respectively, pad the image with black colour (0, 0, 0):

   a) If $a < 0$, pad the image on the left by $p_x = -a$ pixels, set $c = c + a$ and $a = 0$.

   b) If $b < 0$, pad the image on the top by $p_y = -b$ pixels, set $d = d + b$ and $b = 0$.

   c) If $c \geq w$, pad the image on the right by $c - w + 1$ pixels.

   d) If $d \geq h$, pad the image on the right by $d - h + 1$ pixels.

5) Crop I to the rectangle defined by the corners (a, b) and (c, d).

6) Resize the image I to size 120 px × 120 px using the OpenCV function "resize" with bilinear interpolation.

7) Normalize I using mean $\mu=127.5$ and standard deviation $\sigma=128$ as specified in 6.19 to obtain the array A.

8) Encode A as input tensor $T^{[0]}$ of dimensions (1, 3, 120, 120) for the neural network model M.

9) Run a forward pass through the model M object using $T^{[0]}$ as input to obtain an output tensor $T^{[1]}$ of dimension (1, 62).

10) Set $R_1=(R_{1,1},R_{1,2},R_{1,3})=\left(T_1^{[1]},T_2^{[1]},T_3^{[1]}\right)$.

11) Set $R_2=(R_{2,1},R_{2,2},R_{2,3})=\left(T_5^{[1]},T_6^{[1]},T_7^{[1]}\right)$.

12) De-normalize R_1 by mean μ = (3.4926363e-04, 2.5279013e-07, −6.8751979e-07) and standard deviation σ = (1.7632153e-04, 6.7379435e-05, 4.4708489e-04), i.e. set:

$$R_{1,i}=\sigma_i\cdot R_{1,i}+\mu_i$$

13) De-normalize R_2 by mean μ' = (−6.2955132e-07, 5.7572004e-04, −5.0853912e-05) and standard deviation σ' = (1.2313770e-04, 4.4930217e-05, 7.9236706e-05), i.e. set:

$$R_{2,i}=\sigma'_i\cdot R_{2,i}+\mu'_i$$

14) Normalize R_1 and R_2 to unit length.

15) Set $R_3 = (R_{3,1}, R_{3,2}, R_{3,3}) = R_1 \times R_2$, where × denotes the cross product.

16) Compute the yaw, pitch and roll angles ϕ_{yaw}, ϕ_{pitch}, ϕ_{roll} as follows:

a) If $R_{3,1} \leq 0.9975$ and $R_{3,1} \geq -0.9975$, set:

$$\phi_{\text{yaw}} = -\text{atan2}\left(\frac{R_{3,2}}{\cos(\phi_{\text{pitch}})}, \frac{R_{3,3}}{\cos(\phi_{\text{pitch}})}\right),$$

$$\phi_{\text{pitch}} = \arcsin(R_{3,1}),$$

$$\phi_{\text{roll}} = -\text{atan2}\left(\frac{R_{2,1}}{\cos(\phi_{\text{pitch}})}, \frac{R_{1,1}}{\cos(\phi_{\text{pitch}})}\right);$$

b) If $R_{3,1} < -0.9975$, set:

$$\phi_{\text{yaw}} = -\text{atan2}(R_{1,2}, R_{1,3}),$$

$$\phi_{\text{pitch}} = -\frac{\pi}{2},$$

$$\phi_{\text{roll}} = 0;$$

c) If $R_{3,1} > 0.9975$, set:

$$\phi_{\text{yaw}} = \text{atan2}(R_{1,2}, R_{1,3}),$$

$$\phi_{\text{pitch}} = \frac{\pi}{2},$$

$$\phi_{\text{roll}} = 0.$$

17) Output ϕ_{yaw}, ϕ_{pitch}, ϕ_{roll}.

6.12 Conversion of 8-bits-per-channel colour images to luminance

Some quality component computations accept luminance data. The luminance conversion from RGB components depends on the International Color Consortium (ICC) profile of the image. Profiles are used for correct sampling and natural display purposes of the RGB values.

For an image, I, coded in sRGB ICC profile, each of its pixels' RGB data luminance can be recovered by using the following steps.

1) If I has only one colour channel (grey scale image), convert I to a colour image with RGB colour channel order.

2) For an image coded in sRGB ICC profile defined in IEC 61966-2-2, skip to step 5.

3) For an image with no known ICC profile, assume the sRGB profile and skip to step 5.

4) For images with other known RGB ICC profiles, convert to the device-independent sRGB colour space based on the image's RGB ICC profile, and continue to step 5.

5) For each pixel's red, green and blue channel values, R, G, B, normalize and perform gamma inversion using

$R_L = \text{ColourConvert}(R/255)$ $G_L = \text{ColourConvert}(G/255)$ $B_L = \text{ColourConvert}(B/255)$

where

$$\mathrm{ColourConvert}(\mathrm{V}) = \begin{cases} ((\mathrm{V}+0.055)/1.055)^{2.4} & \text{if } \mathrm{V} > 0.04045 \\ \mathrm{V}/12.92 & \text{if } \mathrm{V} \le 0.04045 \end{cases},$$

then compute a normalized luminance:

$$Y = 0.2126 R_\mathrm{L} + 0.7152 G_\mathrm{L} + 0.0722 B_\mathrm{L}$$

producing a normalized value on [0;0.999]. Finally, scale the normalized luminance to an integer in the interval [0;255] by setting $Y' = \lfloor 256\,Y \rfloor$.

6.13 Conversion of 8-bits-per-channel colour images to CIELAB space

CIE 1976 L*a*b*[5] is the three-dimensional, approximately uniform colour space produced by plotting in rectangular coordinates L*, a*, b* quantities, i.e. lightness, a chroma and b chroma values.

For each pixel:

1) Normalize and perform gamma inversion of the red, green and blue channel values, R, G, B, using

R_L = ColourConvert(R/255) G_L = ColourConvert(G/255) B_L = ColourConvert(B/255)

where

$$\mathrm{ColourConvert}(\mathrm{V}) = \begin{cases} ((\mathrm{V}+0.055)/1.055)^{2.4} & \text{if } \mathrm{V} > 0.04045 \\ \mathrm{V}/12.92 & \text{if } \mathrm{V} \le 0.04045 \end{cases}.$$

2) Transform to XYZ space:

$$X = 0.43605 R_\mathrm{L} + 0.38508 G_\mathrm{L} + 0.14309 B_\mathrm{L}$$
$$Y = 0.22249 R_\mathrm{L} + 0.71689 G_\mathrm{L} + 0.06062 B_\mathrm{L}$$
$$Z = 0.01393 R_\mathrm{L} + 0.09710 G_\mathrm{L} + 0.71419 B_\mathrm{L}$$

3) Convert XYZ to Lab by normalizing by CIE standard illuminant D50 which simulates warm daylight at sunrise or sunset with correlated colour temperature of 5003K (also known as horizon light):

$$x_r = \frac{X}{0.964221} \qquad y_r = Y \qquad z_r = \frac{Z}{0.825211}$$

5. Flatten the values:

$$F_x = cubic(x_r) \qquad F_y = cubic(y_r) \qquad F_z = cubic(z_r)$$

where

$$cubic(x) = \begin{cases} (kx+16)/116 & \text{if } x \le \varepsilon \\ \sqrt[3]{x} & \text{if } x > \varepsilon \end{cases} \quad \text{and} \quad \varepsilon = \frac{216}{24389} \qquad k = \frac{24389}{27}$$

6. The CIELAB values are:

$$L^* = 116\,F_y - 16 \qquad a^* = 500\left(F_x - F_y\right) \qquad b^* = 200\left(F_y - F_z\right)$$

6.14 Handling of greyscale images

Implementations of this document should function on both colour and greyscale images.

NOTE Greyscale images are achromatic colour images devoid of hue. The RGB scale is calibrated so that when R, G and B values are equal, the colour is a shade of grey without any bias towards red, green or blue hue. In other words, when in any profile R, G and B values are all 0, then the hue = 0 in the CIELAB colour space, i.e. respective CIELAB a^* and b^* values are 0.

6.15 Luminance histogram

Given an image I in RGB colour channel order and, optionally, a segmentation map S of same width and height, its histogram $h_0, \ldots, h_M$ ($M = 255$) is computed as follows:

1) If no segmentation map S is given, set S to the map covering the complete image region of I, i.e. an array of ones with the same width and height as I.

2) Recover the image luminance C of I as specified in 6.12.

3) Compute, for all $0 \le i \le M$, the number n_i of pixels (a,b) with $S_{\mathrm{a,b}} = 1$ and $C'_{\mathrm{a,b}} = 1$.

4) Compute the number N of pixels segmented by S as $N = \sum_i n_i$.

5) Compute $h_i = \dfrac{n_i}{N}$ for all $0 \le i \le M$.

6) Output $h_0, \ldots, h_M$.

6.16 Entropy

Given the 6.15 luminance histogram $h_0, \ldots, h_\mathrm{M}$, the entropy H should be computed as:

$$H = -\sum_{i=0,\, h_i \ne 0}^{M} h_i \log_2 h_i$$

6.17 Expressing binary quantities as continuous values

Some ISO/IEC 39794-5:2019 requirements include a numerical specification such that a Level 3 conformance test on an image would yield a Boolean result. Conformance to such requirements can be softened to produce a continuous quality value using, for example the sigmoid function:

$$\mathrm{SIGMOID}(x, x_o, w) = (1 + \exp((x_0 - x)/w))^{-1}$$

where x_0 is a location parameter at which the function has value 0.5, and w is a scale parameter indicating the width of the region in which the function transitions from ε to $1 - \varepsilon$.

NOTE 1 The non-linear mapping can be useful to avoid assigning much higher quality values to very good images vs. simply good images, and for assigning roughly equal quality values for images with poor and very poor conformance to a requirement.

NOTE 2 An approximately linear transformation to [0;1] can be effected by choosing x_0 and w to correspond to the centre part of the sigmoid function.

6.18 Representation and arithmetic of real and integer numbers

This document specifies algorithms performing computations typically based on real numbers. The document intentionally does not specify the use of mandatory explicit data types since implementations conformant to this document shall work on a variety of operating systems and platforms.

Algorithms compliant to this document may use other data types than 64-bit wide floating point data types if the quality components reproduce the target quality measure values in Table A.3 and Table A.4 of the reference implementation OFIQ.

A real non-negative number, x, may be converted to an integer number using:

$$\mathrm{ROUND}(x) = \lfloor x + ½ \rfloor$$

where $\lfloor \cdot \rfloor$ denotes the floor function that converts a real number, x, to the largest integer smaller than or equal to x. A negative number x, may be converted to an integer number using:

$$\mathrm{ROUND}(x) = \lceil x - ½ \rceil$$

where $\lceil \cdot \rceil$ denotes the ceiling function that converts a real number, x, to the smallest integer larger than or equal to x.

NOTE 1 Real numbers can be represented using a data format with sufficient precision, i.e. for most platforms, typically a 64-bit floating-point data type (i.e. double type).

NOTE 2 Integers can be encoded using an integer data type encoded by a sufficient number of bits. In most cases, for example, for encoding a pixel coordinate value, a 32-bit wide signed integer is sufficient.

NOTE 3 Rounding is following the behaviour of the C++ function std::round(), which implements IEEE 754 “Round to nearest, ties away from zero” mode.[45]

6.19 Normalization of image colour values

For image processing using CNNs, the colour values of the input image need to be normalized with respect to the mean, μ, and standard deviation, σ. The specific values of μ and σ are specific to the CNN. Since the input image has three colour channels, both μ and σ are either 3-dimensional vectors or scalars that are replicated to obtain 3-dimensional vectors.

Normalization of image colour values takes as input an image I with 3 colour channels and the values μ and σ, and is performed using the following procedure.

1) If μ is a scalar, set $\mu_1 = \mu_2 = \mu_3 = \mu$.
2) If σ is a scalar, set $\sigma_1 = \sigma_2 = \sigma_3 = \sigma$.
3) Set $A_{i,j,k} = (I_{i,j,k} - \mu_k)\, \sigma_k^{-1}$ for all pixel coordinates i, j, and for all colour channels, k.
4) Output A.

7 Quality measures

7.1 General

An implementation conforming to this document shall include the quality measure computations defined in 7.2 to 7.4, as required by Table 3.

In cases where an implementation of a quality assessment algorithm does not output a value for a quality measure, for example due to the failure to localize both eyes, the error code shall be set to the value required in ISO/IEC 29794-1, which is the abstract value failureToAssess.

If a quality component is listed as optional in Clause 5, it may be omitted.

While the native quality measures are generally computed and reported as real numbers, if included in one of the biometric data interchange formats specified in the ISO/IEC 19794 series and the ISO/IEC 39794 series, they shall be reported as non-dimensional numbers between 0 and 100 that are encoded as integers, as specified in ISO/IEC 29794-1.

NOTE 1 The performance of the quality assessment algorithms defined here has been tested with regards to being predictive for recognition accuracy. Annex E references OFIQ test reports.

NOTE 2 Annex F gives guidance on how quality can be used to improve captured images during a capture session.

NOTE 3 All the results from the quality measure computations constitute a quality vector. The components of the vector will generally be correlated – that is, over a large set of images, the correlation matrix will not be diagonal.

NOTE 4 A test of a face image quality assessment algorithm could demonstrate dependence of quality components given in Clause 7 on demographic factors, such as skin tone or age.

NOTE 5 When an implementation of this document is called with an image containing multiple faces, the quality of the largest face is assessed. If an application seeks to assess the quality of other faces, or multiple faces, then the calling application could prepare sub-images by cropping appropriate areas and make multiple calls to the QAA.

7.2 Quality score (unified)

7.2.1 Description

An implementation shall expose a function that accepts an input image and outputs a quality score. The implementation shall accept both greyscale and colour images. Given an image I, an implementation, F, shall produce a scalar quality score, $Q = \mathrm{F}(I)$. The function, F, should implement the recognition-outcome prediction requirements of ISO/IEC 29794-1:2024, 11.2 and Clause B.2 of this document.

NOTE These provisions would also be satisfied by a function present in a face recognition software development kit which produces a face recognition template and a quality score.

Four examples are shown in Figure 7. The progression, from left to right, implies that images closer to canonical face images have higher quality scores.

QS = 95	QS = 85	QS = 62	QS = 42
a	b	c	d

Figure 7 — Four face images with example image quality values

7.2.2 Computation of the native quality measure

The reference implementation[58] uses a model converted from the original PyTorch[54] which is based on certain publications.[55], [56] The algorithm takes as input the image I in BGR colour channel order (8 bits per channel), output by the alignment algorithm in 6.7, and performs the following steps:

1) Resize I to 192 px × 192 px using bilinear interpolation.

2) Crop I by 40 pixels from left and right, by 33 pixels from the top and by 47 pixels from the bottom.

3) Using mean $\mu = 0$ and standard deviation $\sigma = 255$, normalize the cropped image using 6.19.

4) Encode A as input tensor T of dimensions (1, 1, 112, 112).

5) Run a forward pass through the model M using T as input to obtain an output tensor T' of dimensions (1, 1, 512).

6) Cast T' to a vector V of length 512.

7) Set q to the Euclidean norm of V, i.e. $q = \|V\|_2$.

8) Output q.

7.2.3 Mapping the computation result to the target range of the quality component

The native quality measure q shall be mapped to a quality score using:

Q = ROUND(100 SIGMOID(q, 23, 2.6))

7.3 Capture-related quality components

7.3.1 General

The quality components of the subsequent subclauses are related to an imaging system and acquisition environment.

These quality components have concrete meaning that they can be acted upon for re-capture. Therefore, they can be used to monitor a system or to assess quality requirements for specific use-cases.

EXAMPLE If a capture system produces face images with a low mean value for the sharpness quality component, an operator could check the focus of the system. If a system produces images with a low mean value for the illumination uniformity quality component, an operator could improve the lighting environment.

7.3.2 Background uniformity

7.3.2.1 Description

The background is any surface or scene that is behind the biometric capture subject. Uniform backgrounds are sometimes specified to improve face detection. Three background checks form the basis for the background quality computation i.e. uniformity, texturing and white balance.

This quality measure refers to the requirements in ISO/IEC 19794-5:2011, 7.2.6 (relevant for UC2 as specified in Reference [64]) and ISO/IEC 39794-5:2019, D.1.4.2.5 (relevant for UC1 specified Reference [65]).

7.3.2.2 Computation of the native quality measure

The algorithm takes as input the following:

1) Image I of dimension (616, 616) and the affine transformation matrix M output by the alignment algorithm in 6.7.

2) Segmentation map S of dimension (400,400) computed with the face parsing algorithm in 6.8.

3) The dimensions (w,h) of the original face image (prior to alignment).

The computation proceeds as follows.

1) Create a completely black (0) grey scale image A of dimensions (w,h).

2) Apply the transformation M *to* A and crop and/or pad the image to dimension 616 px × 616 px to obtain a mask P for the regions that were padded during alignment. For padding use white colour (255).

3) Crop both I and P by 62 pixels from both sides and by 210 pixels from the bottom to adjust the image region of I and P to that covered by the segmentation map S.

4) Resize both I and P to the width of 354 pixels and height of 292 pixels, where I is resized using bilinear interpolation and P is resized using nearest integer interpolation.

5) Crop the segmentation map S by 23 pixels from both sides and by 108 pixels from the bottom.

6) Compute the background mask B with $B_{ij} = 1$, if $S_{ij} = 0$ and $P_{ij} = 0$, and $B_{ij} = 0$ otherwise.

7) Apply to B the OpenCV function erode with kernel size 4.

8) If $B_{ij} = 0$ for all (i,j), output failureToAssess and terminate the algorithm.

9) Recover the luminance image L from the RGB data of I as specified in 6.12.

10) Apply horizontal and vertical Sobel filters with kernel size 3, S_3^x and S_3^y respectively, to L.

11) Compute the array, G, of the magnitudes of the gradients of L as:

$$G_{i,j} = \sqrt{S_3^x(L)_{i,j}^{\ 2} + S_3^y(L)_{i,j}^{\ 2}}\ .$$

12) Compute the mean length of the gradients on the background:

$$\mu_l = \left(\sum_{(i.j):B_{i,j}=1} G_{i,j}\right) / \sum_{(i.j)} B_{i,j}$$

13) Output μ_l .

7.3.2.3 Mapping the computation result to the target range of the quality component

The native quality measure (i.e. the mean length of the gradient μ_l) has smaller-is-better semantics. The native quality measure shall be mapped to a quality component value using:

Q = min(100, max(0, ROUND(190 (1 - SIGMOID(μ_l, 10, 100)))))

7.3.3 Illumination uniformity

7.3.3.1 Description

Measure the difference in illumination on the left and right side of the face.

This quality measure refers to the requirements in ISO/IEC 19794-5:2011, 7.2.7 and 7.2.8 (relevant for UC2 as specified in Reference [64]) and of ISO/IEC 39794-5:2019, D.1.4.2.5, D.1.4.2.6 and D.1.4.5.2 (relevant for UC1 specified in Reference [65]).

For UC1, ISO/IEC 39794-5:2019, D.1.4.2.5 requires absence of asymmetric shadows. Further, ISO/IEC 39794-5:2019, D.1.4.2.6 on lighting requires that face portraits have adequate and uniform illumination. Also, ISO/IEC 39794-5:2019, D.1.4.5.2 on children below one year requires shadows of an assistant to not be visible on the face.

NOTE Face luminance as measured from an image will depend on incident light and skin properties, including dermatological conditions and possible presence of tattoos. Measured left-right symmetry will depend on actual uniformity of the incident light and on head orientation.

7.3.3.2 Computation of the native quality measure

The algorithm makes use of the intensity measurement zones L and R as described in ISO/IEC 39794-5:2019, Figure D.14 and its accompanying lighting-related text in ISO/IEC 39794-5:2019, D.1.4.2.6. It takes as input the aligned image and the transformed landmarks output by the algorithm in 6.7.

1) From the transformed face landmarks, compute the centres (X_L,Y_L) and (X_R,Y_R) of the left and right eyes in the aligned image as output by the algorithm in 6.10.

2) Set $d_{IED} = ||(X_L,Y_L)-(X_R,Y_R)||_2$ to the Euclidean distance of the two eye centres.

3) Compute the midpoint between the eyes, $M = ((X_L, Y_L) + (X_R, Y_R)) / 2$.

4) Compute the mouth midpoint as $N = (T_{90} + T_{94}) / 2$

5) Compute the eye mouth distance, $d_{EMD} = ||M - N||_2$.

6) Set L' to be the square with corners $(X_L - 0.3\ d_{IED}, Y_L + 0.5\ d_{EMD})$ and $(X_L, Y_L + 0.5\ d_{EMD} + 0.3\ d_{IED})$.

7) Set R' to be the square with corners $(X_R, Y_R + 0.5\ d_{EMD})$ and $(X_R + 0.3\ d_{IED}, Y_R + 0.5\ d_{EMD} + 0.3\ d_{IED})$.

8) Recover the image luminance from the RGB data in L' and R', respectively, as specified in 6.12.

9) Compute luminance histograms (normalized as described in 6.15) of L' and R'.

10) Compute the histogram intersection of the luminance histograms $p_0, \ldots, p_{255}$ and $q_0, \ldots, q_{255}$ of the left and right halves of the face bounding box:

$$D = \sum_{i=0}^{255} \min(p_i, q_i)$$

11) Output D.

NOTE 1 The more symmetric the intensity measurement zones, the higher D. It is understood that most faces are asymmetric such that this partitioning will only be approximate.

NOTE 2 Large yaw and pitch angles have a significant impact on the computation of the measurement zones L' and R' in Step 3 and in Step 4. Operators could inspect large deviations of head pose angles first before assessing defects of the illumination uniformity.

Specular and diffuse reflection components within specific face image regions can be separated using the Dichromatic Reflection Model.[27]

7.3.3.3 Mapping the computation result to the target range of the quality component

The native quality measure (i.e. histogram difference aspect, D) has smaller-is-worse semantics, with an ideal value of 1. It shall be mapped to a quality component value using:

$$Q = \mathrm{ROUND}(100\ D^{0.3})$$

7.3.4 Moments of the luminance distribution

7.3.4.1 General

The brightness of an image can be investigated by considering the histogram of the luminance values, e.g. the luminance in the YC_BC_R colour system. The histogram h_i of luminance values, $0 \le i \le M$, where M is the maximum possible luminance value (255 in 8-bit representations), can be considered to be a probability density function that can be characterized by its statistical moments (see 6.15).

This quality measure refers to the requirements in ISO/IEC 19794-5:2011, 7.2.8 (relevant for UC2 as specified in Reference [64]) and ISO/IEC 39794-5:2019, D.1.2.4.6 (relevant for UC1 specified in Reference [65]).

For UC1, ISO/IEC 39794-5:2019, D.1.4.2.6 requires face portraits to have adequate and uniform illumination.

7.3.4.2 Luminance mean

7.3.4.2.1 Description

As too high or too low brightness can reduce recognition accuracy, this quality component computes the mean luminance.

7.3.4.2.2 Computation of the native quality measure

The algorithm takes as input the image *I* and the landmarked region *S* output by the algorithm in 6.6.

1) Recover the image luminance from RGB data of *I* as specified in 6.12.

2) Compute the luminance histogram $h_0, \cdots, h_M$ (M=255) in *S* as specified in 6.15.

3) Compute the mean of the luminance on *S* as:

$$\mu = \sum_{i=0}^{M} \frac{h_i}{M}$$

4) Output μ.

7.3.4.2.3 Mapping the computation result to the target range of the quality component

The native quality measure value μ shall be mapped to a quality component value using:

$$Q = \lceil 100\, SIGMOID(\mu, 0.2, 0.05)(1 - SIGMOID(\mu, 0.8, 0.05)\rceil$$

This penalizes very dark and very bright images but assigns high quality to a wide central range. The function is symmetric about $\mu = 0.5$.

7.3.4.3 Luminance variance

7.3.4.3.1 Description

The second moment relates to contrast of the image. Sufficiently high or low variance will undermine recognition.

7.3.4.3.2 Computation of the native quality measure

The algorithm takes as input the image *I* and the landmarked region *S* output by the algorithm in 6.6.

1) Compute the luminance histogram $h_0, \cdots, h_M$ (M=255) in *S* as specified in 6.15.

2) Compute the luminance variance on *S* as:

$$VAR(h_0, \cdots, h_M) = \sum_{i=0}^{M} h_i \left(\frac{i}{M} - \bar{h} \right)^2 .$$

where $\bar{h}$ is the value output by the algorithm in 7.3.4.2.2.

3) Output *VAR.*

7.3.4.3.3 Mapping the computation result to the target range of the quality component

The native quality measure (i.e. the variance, *VAR*) takes on low values when the luminance histogram is concentrated around the mean value. The native quality measure shall be mapped to a quality component value using:

$$Q = \text{ROUND}\left(100 \sin\left(\frac{60\, VAR}{60\, VAR + 1} \pi \right) \right)$$

7.3.5 Under-exposure prevention

7.3.5.1 Description

The representation of a face is too dark if it has a high proportion of pixels that have a low luminance value.

This quality measure refers to the requirements in ISO/IEC 19794-5:2011, 7.3.2 (relevant for UC2 as specified in Reference [64]) and ISO/IEC 39794-5:2019, D.1.4.2.7 (relevant for UC1 as specified in Reference [65]).

For UC1, ISO/IEC 39794-5:2019, D.1.4.2.7 requires contrast gradations in textures to be clearly visible and requires face portrait to have appropriate brightness and good contrast.

NOTE A biometric capture subject invariant lighting set up can result in under- or over- exposure of the face depending upon skin pigmentation of the biometric capture subject.

7.3.5.2 Computation of the native quality measure

The algorithm takes as input the aligned image *I* as output by the algorithm in 6.7. Furthermore, the algorithm takes as input the face occlusion segmentation *O* output by the algorithm in 6.9 and the landmarked region *S* output by the algorithm in 6.6.

1) Compute the non-occluded landmarked region *S'* as the intersection of *S* and the regions not marked occluded in *O*.

2) If *S'* is empty, output failureToAssess and stop, otherwise continue to the next step.

3) Recover the image luminance from RGB representation of *I* as specified in 6.12.

4) Compute the luminance histogram $h_0, \ldots, h_{255}$ in the non-occluded landmarked region *S'* as specified in 6.15.

5) From the luminance histogram, compute the proportion of pixels for which luminance is on range [0;25], using:

$$v = \sum_{i=0}^{25} h_i$$

6) Output v.

7.3.5.3 Mapping the computation result to the target range of the quality component

The native quality measure, *v,* has larger-is-worse semantics. It shall be mapped to a quality component value using:

$$Q = \mathrm{MIN}(100, \mathrm{MAX}(0, \mathrm{ROUND}(120\ (0.832 - \mathrm{SIGMOID}(v, 0.92, 0.05)))))$$

7.3.6 Over-exposure prevention

7.3.6.1 Description

A face is light if it has a high proportion of pixels that have a high luminance value i.e. hot spots.

This quality measure refers to the requirements in ISO/IEC 19794-5:2011, 7.3.2 (relevant for UC2 as specified in Reference [64]) and ISO/IEC 39794-5:2019, D.1.4.2.7 (relevant for UC1 specified in Reference [65]).

For UC1, ISO/IEC 39794-5:2019, D.1.4.2.7 requires contrast gradations in textures to be clearly visible and requires face portraits to have appropriate brightness and good contrast.

NOTE A biometric capture subject invariant lighting set up can result in under- or over-exposure of the face depending upon skin pigmentation of the biometric capture subject.

7.3.6.2 Computation of the native quality measure

The algorithm takes as input the aligned image I as output by the algorithm in 6.7. Furthermore, the algorithm takes as input the face occlusion segmentation O output by the algorithm in 6.9 and the landmarked region S output by the algorithm in 6.6.

1) Compute the non-occluded landmarked region S' as the intersection of S and the regions not marked occluded in O.

2) If S' is empty, output failureToAssess and terminate; otherwise continue with the next step.

3) Recover the image luminance from RGB representation of I as specified in 6.12.

4) Compute the luminance histogram $h_0,\ldots,h_{255}$ in the non-occluded landmarked region S' as specified in 6.15.

5) From the luminance histogram, compute the proportion of pixels for which luminance is on range [247;255] using:

$$v = \sum_{i=247}^{255} h_i$$

6) Output v.

7.3.6.3 Mapping the computation result to the target range of the quality component

The native quality measure, v, has larger-is-worse semantics. It shall be mapped to a quality component value using:

$$Q = \mathrm{ROUND}(1 / (v + 0.01))$$

7.3.7 Dynamic range

7.3.7.1 Description

This quality measure refers to the requirements in ISO/IEC 19794-5:2011, 7.3.5 (relevant for UC2 as specified in Reference [64]) and ISO/IEC 39794-5:2019, D.1.4.2.8 (relevant for UC1 as specified in Reference [65]).

For UC1, ISO/IEC 39794-5:2019, D.1.4.2.8 recommends the dynamic range of all colour channels of a face image have at least 50 % of intensity variation in the facial region of the image.

7.3.7.2 Computation of the native quality measure

The algorithm takes as input the aligned image I and the aligned landmark as output by the algorithm in 6.7.

1) Compute the landmarked region S using the aligned landmark by applying the algorithm in 6.6 with the aligned landmarks as input.

2) Recover the image luminance from RGB representation of I as specified in 6.12.

3) Compute the luminance histogram restricted to the landmarked region S using the algorithm specified in 6.15.

4) Compute the entropy, H, of the luminance histogram using the algorithm specified in 6.16.

5) Output H.

NOTE ISO/IEC 39794-5:2019, D.2.4.4 indicates $H \geq 7$, otherwise the image is not correctly exposed. This recommendation can require camera or video digitizer settings to be changed for each individual biometric capture subject to account for any differences in skin tone.

7.3.7.3 Mapping the computation result to the target range of the quality component

The native quality measure, H, has larger-is-better semantics. It shall be mapped to a quality component value using:

Q = MIN(100, ROUND(12.5 H))

7.3.8 Sharpness

7.3.8.1 Description

Focus can be measured by edge-detecting features like Sobel and Laplace filters, but also by measuring the difference between the image and a mean-filtered version of the image.

This quality component can be used to efficiently choose the better focused face portrait among several face samples of the same biometric capture subject.

This quality measure refers to the requirements in ISO/IEC 19794-5:2011, 7.3.8 (relevant for UC2 as specified in Reference [64]) and ISO/IEC 39794-5:2019, D.1.4.2.4 (relevant for UC1 as specified in Reference [65]).

For UC1, ISO/IEC 39794-5:2019, D.1.4.2.4 requires images to have sufficient focus and depth of field to maintain the required level of details.

NOTE 1 This quality measure generically captures various degradations that remove high spatial frequency information including optical blur, loss of focus and atmospheric turbulence.

NOTE 2 The sharpness score can be artificially inflated by JPEG compression and variable sensor noise, such as produced by sensor-gain-based exposure control.

7.3.8.2 Computation of the native quality measure

The reference implementation uses a pre-trained random forest classifier available from Reference [58]. It takes as input the image I in BGR colour channel order with 8 bits per channel and the landmarked region M output by the algorithm in 6.6, and performs the following steps.

1) If the landmarked region protrudes the image region, restrict it to its intersection with the image region.

2) Crop I to the minimal upright bounding box of the landmarked region.

3) Convert I to grey scale.

4) For each of the kernel sizes s = 1, 3, 5, 7, 9, compute the isotropic Sobel filter S_s^{xy} on I, and compute the arithmetic mean μ_s^S and standard deviation σ_s^S of the absolute values of the filter outputs on the landmarked region.

5) For each of the kernel sizes s = 3, 5, 7, apply a mean filter to I, compute the difference of the resulting image from I, and compute the arithmetic mean μ_s^M and standard deviation σ_s^M of the difference on the landmarked region.

6) Apply a Gaussian blur filter with (3 × 3) kernel on I to remove noise.

7) For each of the kernel sizes s = 1, 3, 5, 7, 9, compute the Laplace filter L_s^{xy} on I, and compute the arithmetic mean μ_s^L and standard deviation σ_s^L of the absolute values of the filter outputs on the landmarked region.

8) Compute a 26-dimensional feature vector:

$$f = \left(\left(\mu_i^S, \sigma_i^S \right)_{i=1,3,5,7,9}, \left(\mu_i^M, \sigma_i^M \right)_{i=3,5,7}, \left(\mu_i^L, \sigma_i^L \right)_{i=1,3,5,7,9} \right)$$

9) Feed the feature vector f into the random forest classifier model to obtain a score x.

10) Return x.

7.3.8.3 Mapping the computation result to the target range of the quality component

The native quality measure, x, has smaller-is-worse semantics. It shall be mapped to a quality component value using:

Q = ROUND(115 SIGMOID(x, −20, 15) − 14)

7.3.9 No compression artefacts

7.3.9.1 Description

Lossy compression, including the common ISO/IEC 10918 series JPEG method, can undermine face recognition accuracy because it removes information from an image, and can introduce artefacts. The subclause computes a measure of the visibility of compression artefacts in the face region.

This quality measure refers to the requirements in ISO/IEC 19794-5:2011, 5.7.3 and 6.2 (relevant for UC2 as specified in Reference [64]) and in ISO/IEC 39794-5:2019, D.1.4.2.4, D.1.5.2 and D.1.5.5 (relevant for UC1 as specified in Reference [65]).

For UC1, ISO/IEC 39794-5:2019, D.1.5.5, requires that captured face portraits of a subject do not sacrifice image quality by overly compressing the image keeping the compression ratio at or below 15:1.

NOTE Compression is applied at capture time, under control of the system owner, so does not need to be measured in a capture loop. It can be appropriate to measure compression ratios when surveying of images in legacy databases, or in ID documents from various issuers, or over different collection sites.

7.3.9.2 Computation of the native quality measure

The reference implementation uses a pre-trained model available from Reference [58]. The algorithm takes as input the aligned image I in RGB colour channel order (8 bits per channel) output by the algorithm in 6.7.

1) Crop I by 184 pixels from all sides.

2) Normalize I with mean μ = (123.7, 116.3, 103.5) and standard deviation σ = (58.4, 57.1, 57.4) as specified in 6.19.

3) Reshape I as input tensor T of dimensions (1, 3, 248, 248) for the neural network model.

4) Run a forward pass through the CNN model using T as input to obtain a single output value tensor T' of dimension (1,1).

5) Output $q = T'_{1,1}$.

NOTE 1 Some compression algorithms can produce a larger output than input for some unusual inputs, e.g. already compressed data.

NOTE 2 More information on the impact of compression on face image sample quality and recognition accuracy can be found in References [28] and [29].

NOTE 3 The CNN has been trained to handle two cases: first, where compression is applied after cropping, scaling and rotation; and second, where scaling or rotation has been applied after compression.

NOTE 4 The CNN has been trained to detect both JPEG and JPEG 2000 artefacts.

7.3.9.3 Mapping the computation result to the target range of the quality component

The native quality measure, q, has lower-is-worse semantics. It shall be mapped to a quality component value using:

Q = MAX(0, MIN(100, ROUND(103 SIGMOID(q, 0.3308, 0.092) - 0.0278))).

7.3.10 Natural colour

7.3.10.1 Description

This quality measure refers to the requirements in ISO/IEC 19794-5:2011, 5.7.9, 7.3.4, 7.3.5, 7.4.2.1 and 7.4.4, frontal image type (relevant for UC2 as specified in Reference [64]) and in ISO/IEC 39794-5:2019, D.1.4.2.4 and D.1.4.2.9 (relevant for UC1 as specified in Reference [65]).

For UC1, ISO/IEC 39794-5:2019, D.1.4.2.9 discusses unnatural skin tones based on consideration of a* b* from the CIELAB L*a*b* colour space.[5]

NOTE 1 In order to achieve high fidelity skin tones, the white balance setting is important for photography. In ISO/IEC 39794-5:2019, E.3.3, the reference to colour balance has the connotation of white balance. White balance is measured indirectly using skin tone as shown in this unnatural colour computation.

NOTE 2 Skin tones with a* b* values outside an expected range can result for reasons other than incorrect white balance settings including from medical conditions as well as facial alterations (e.g. face tattoos).

NOTE 3 ISO/IEC 39794-5:2019, E.3 and ICAO only accept images for which a* and b* are positive.

7.3.10.2 Computation of the native quality measure

Typical skin colour *L*a*b** values fit between 5-25 for *a** and 5-35 for *b**.[22],[23] To cover all ethnic groups the limits are set wider than shown in research papers describing the skin tone distributions.[24] Based on these limit values the skin tone plateau mean values are 15 for *a** and 20 for *b**. A large facial tattoo or painting should produce low unnatural skin tone quality value thus measuring the utility of a facial image for face recognition[25]. The wide skin tone plateau ensures diverse human skin tones are reported as natural.

The algorithm takes as input the aligned image and the transformed landmarks T_0 ... T_{97} output by the algorithm in 6.7.

1) If an image has a single luminance channel and no chrominance channels, set D = 0 and stop.

2) If all pixel values of all three colour channels are identical, set D = 0 and stop.

3) From the transformed face landmarks, compute the centres (X_L,Y_L) and (X_R,Y_R) of the left and right eyes in the aligned image as output by the algorithm in 6.10.

4) Set $d_{IED} = || (X_L,Y_L) - (X_R,Y_R) ||_2$ i.e. the Euclidean distance between the two eye centres.

5) Compute the midpoint between the eyes, $M = ((X_L, Y_L) + (X_R,Y_R)) / 2$.

6) Compute the mouth midpoint as $N = (T_{90} + T_{94}) / 2$.

7) Compute the eye mouth distance $d_{EMD} = ||M - N||_2$.

8) Set L' to be the square with corners $(X_L - 0.3\ d_{IED}, Y_L + 0.5\ d_{EMD})$ and $(X_L, Y_L + 0.5\ d_{EMD} + 0.3\ d_{IED})$.

9) Set R' to be the square with corners $(X_R, Y_R + 0.5\ d_{EMD})$ and $(X_R + 0.3\ d_{IED}, Y_R + 0.5\ d_{EMD} + 0.3\ d_{IED})$.

10) Using the left and right regions specified in ISO/IEC 39794-5:2019, D.1.4.2.6 and Figure D.14, compute, for each channel R, G and B, the mean value of all pixels in the L' and R' regions.

11) Convert the RGB value to CIELAB colour space using the procedure in 6.13 and retain *a** and *b**.

12) If either a^* or b^* or both are less than zero then $D = 100$ and stop.

10) Combine a^* and b^* values using the following formula:

$$D = \sqrt{\max\left(\max\left(0, 5 - a^*\right), \max\left(0, a^* - 25\right)\right)^2 + \max\left(\max\left(0, 5 - b^*\right), \max\left(0, b^* - 35\right)\right)^2}$$

where D is a measure of distance in (a^*, b^*) space from the skin tone plateau.

NOTE 1 Appropriate colour calibration for a face in an image can be achieved using colour calibration targets with known spectral properties. Colour targets that include a range of colours corresponding to human skin tones ensure more accurate calibration within that range. A pre-computed calibration could be applied to images from a fixed camera and lighting environment, or a colour target can be acquired along with the face image when devices or lighting environments vary.

NOTE 2 Computing the mean RGB values for all pixels is faster than computing CIELAB values for all pixels then taking the mean CIELAB values. Penalizing unnatural colour using the wide plateau will generally provide sufficient margin for errors caused by the image noise, minor skin defects and white balance errors when skin tone is converted to CIELAB a^*b^* value.

NOTE 3 This quality component only evaluates if the skin colour is in a range of plausible human skin colour. It gives no information about whether the face portrait has fidelity and is a faithful representation of the capture subject.

NOTE 4 Large yaw and pitch angles have a significant impact on the computation of the measurement zones L and R in step 4 and in step 5. Operators could inspect large deviations of head pose angles first before assessing natural colour.

7.3.10.3 Mapping the computation result to the target range of the quality component

The native quality measure, D, indicates the un-naturalness of the skin tone, with lower-is-better semantics. It shall be mapped to the natural colour quality component value using:

$Q = \mathrm{ROUND}(200(1 - \mathrm{SIGMOID}(D, 0, 10)))$

7.4 Subject-related quality components

7.4.1 General

The values of the quality components of 7.4.2 to 7.4.13 are related to the subject behaviour and depend on the human factors design of the data capture subsystem.

NOTE If a subject-related quality component is evaluated as low during capture, feedback can be provided to the biometric capture subject so that they can act to improve the acquisition quality.

7.4.2 Single face present

7.4.2.1 Description

This quality measure refers to the requirements in ISO/IEC 19794-5:2011, 7.2.4 (relevant for UC2 as specified in Reference [64]) and ISO/IEC 39794-5:2019, D.1.4.5.2 (relevant for UC1 as specified in Reference [65]).

For UC1, ISO/IEC 39794-5:2019, Clause D.1 requires images to contain exactly one face. However, extraneous additional faces can occur in the background, in wall-mounted posters, or on clothing. A detector can find such faces and pass them to a recognition module incorrectly, causing, for example, a false rejection.

7.4.2.2 Computation of the native quality measure

The algorithm takes as input the image I.

1) Find the faces in I using steps 1 – 10 of the procedure in 6.4.

2) Execute the following computation:

   a) If zero faces are detected output "failureToAssess" and terminate.

   b) If one face is detected, output the face unicity as $f = 0$ and terminate.

   c) Otherwise, for each face detected, compute the face box area as described in Step 10 of 6.4.

   d) Find the area of the largest face box, A_1, and the area of the second largest face box, A_2.

   e) Output the face unicity, $f = A_2/A_1$, and terminate the algorithm.

7.4.2.3 Mapping the computation result to the target range of the quality component

The native quality measure face unicity, f, shall be mapped to a quality component value using:

$Q = \text{ROUND}(100(1\text{-} f))$

7.4.3 Eyes open

7.4.3.1 Description

This quality measure refers to the requirements in ISO/IEC 19794-5:2011, 5.5.7, frontal image type (relevant for UC2 as specified in Reference [64]) and in ISO/IEC 39794-5:2019, D.1.4.3.3 (relevant for UC1 as specified in Reference [65]).

For UC1, ISO/IEC 39794-5:2019, D.1.4.3.3, requires both eyes to be open naturally, but not forced wide-open, and also that pupils and irises, including iris colour, are completely visible (although there may be exceptions due to ethnicity or other individually specific reasons).

7.4.3.2 Computation of the native quality measure

The algorithm takes as input the aligned landmarks L_0, ..., L_{97} output by the algorithm in 6.5 and 6.7.

1) Compute the distance T between the midpoint between the eyes and the chin as:

$$T = \left\| \frac{L_{60} + L_{64} + L_{68} + L_{72}}{4} - L_{16} \right\|_2$$

2) Compute the largest distance between the upper and lower eyelids in the subject's left eye as:

$$D_\text{L} = \max\left(\|L_{69} - L_{75}\|_2 , \|L_{70} - L_{74}\|_2 , \|L_{71} - L_{73}\|_2 \right)$$

3) Compute the largest distance between the upper and lower eyelids in the subject's right eye as:

$$D_\text{R} = \max\left(\|L_{61} - L_{67}\|_2 , \|L_{62} - L_{66}\|_2 , \|L_{63} - L_{65}\|_2 \right)$$

4) Compute the palpebral aperture as the smaller of the two as: $D_\text{PAL} = \min(D_\text{L}, D_\text{R})$

5) Compute the eye openness aspect as:

$$\omega = \frac{D_\text{PAL}}{T}$$

NOTE 1 Measurement of small D_PAL values is difficult in lower resolution images. For example, in an image with an inter-eye distance of 75 pixels, corresponding to approximately one pixel per mm, a palpebral aperture of 2 mm would correspond to 2 pixels.

NOTE 2 An inaccurate estimate (e.g. due to low resolution) has significant potential for bias against demographic groups with naturally low D_{PAL}.

NOTE 3 The use of T to normalize the output will provide invariance to head yaw angle, but introduce a dependence on head pitch angle, but this is expected to be smaller than for the yaw.

7.4.3.3 Mapping the computation result to the target range of the quality component

The native quality measure, ω, has larger-is-better semantics. It shall be mapped to a quality component value using:

Q = ROUND(100 SIGMOID(ω, 0.02, 0.01))

7.4.4 Mouth closed

7.4.4.1 Description

This quality measure refers to the requirements of ISO/IEC 19794-5:2011, 5.5.7 (relevant for UC2 as specified in Reference [64]) and ISO/IEC 39794-5:2019, D.1.4.3.2 (relevant for UC1 as specified in Reference [65]).

For UC1, ISO/IEC 39794-5:2019, D.1.4.3.2 requires the mouth to be closed.

7.4.4.2 Computation of the native quality measure

The algorithm takes as input the landmarks L_0, ..., L_{97} output by the algorithm in 6.5.

1) Compute the distance T between the midpoint between the eyes and the chin:

$$T = \left\| \frac{L_{60} + L_{64} + L_{68} + L_{72}}{4} - L_{16} \right\|_2$$

2) Determine the maximum distance of the upper and lower lip:

$$D_{\mathrm{L}} = \max\left(\|L_{89} - L_{95}\|_2, \|L_{90} - L_{94}\|_2, \|L_{91} - L_{93}\|_2 \right)$$

3) Compute the mouth openness aspect:

$$\omega = \frac{D_{\mathrm{L}}}{T}$$

4) Output ω.

7.4.4.3 Mapping the computation result to the target range of the quality component

The native quality measure, ω, has lower-is-better semantics. It shall be mapped to a quality component value using:

Q = ROUND(100(1 – SIGMOID(ω, 0.2, 0.06)))

7.4.5 Eyes visible

7.4.5.1 Description

This quality measure refers to the requirements in ISO/IEC 19794-5:2011, 5.5.7, 7.2.9, 7.2.11, 7.2.12 and 7.2.13, frontal image type (relevant for UC2 as specified in Reference [64]) and in ISO/IEC 39794-5:2019, D1.4.3.3 and D.1.4.3.4 (relevant for UC1 as specified in Reference [65]).

For UC1, the eye visibility text in ISO/IEC 39794-5:2019, D.1.4.3.3 requires both eyes to be opened naturally, and pupils and irises to be completely visible.

One or both eyes can be absent in an image due to occlusion (eyeglasses, hair, scarves), head rotation (yaw, or pitch), or injury.

7.4.5.2 Computation of the native quality measure

The algorithm takes as input the transformed landmarks output by the algorithms in 6.5 and 6.7, the yaw angle of the head pose, and the face occlusion segmentation, S, output by the algorithm in 6.9.

1) From the transformed face landmarks and the yaw angle, compute the inter-eye distance, d_{IED}, using the algorithm in 6.10.

2) Compute the offset distance of the eye visibility zone (EVZ), $V = \lfloor d_{\mathrm{IED}} / 20 \rfloor$

3) Compute the minimum upright (i.e. not rotated) rectangle R_1 containing the landmarks marking the contour of the subject's right eye. The indices of these landmarks are the integers from 60 to 67.

4) Expand R_1 on all sides by V pixels.

5) Compute the minimum upright (i.e. not rotated) rectangle R_2 containing the landmarks marking the contour of the subject's left eye. The indices of these landmarks are the integers from 68 to 75.

6) Expand R_2 on all sides by V pixels.

7) Compute the EVZ, E, as the union of R_1 and R_2, i.e. $E = E_1 \cup E_2$.

8) Determine the fraction, α, of the EVZ that is occluded as $\alpha = |E \cap O|/|E|$, where $O = \{(x, y) | S(x, y) = 0\}$ is the region marked as occluded in the occlusion segmentation map S, and $|X|$ denotes the area of a region X.

9. Output α.

NOTE The EVZ is defined in ISO/IEC 39794-5:2019, D.1.4.3.3.

7.4.5.3 Mapping the computation result to the target range of the quality component

The native quality measure value, α, shall be mapped to a quality component value using:

$$Q = \mathrm{ROUND}(100\,(1 - \alpha))$$

7.4.6 Mouth occlusion prevention

7.4.6.1 Description

This quality measure refers to the requirements in ISO/IEC 19794-5:2011, 5.5.7 (relevant for UC2 as specified in Reference [64]) and ISO/IEC 39794-5:2019, D.1.4.3.5 (relevant for UC1 as specified in Reference [65]).

It is possible for the mouth not to be present in an image due to occlusion (sanitary mask, scarves) or head rotation (yaw, or pitch).

7.4.6.2 Computation of the native quality measure

The algorithm takes as input the transformed landmarks output by the algorithm in 6.7 and the face occlusion segmentation S output by the algorithm in 6.9.

1) Determine the mouth region M as the region enclosed by the landmarks marking the contour of the mouth. The indices of these landmarks are the integers from 76 to 87.

2) Determine the fraction, α, of the mouth region's area that is occluded, i.e. as $\alpha = |M \cap O|/|M|$, where $O = \{(x,y) | S(x,y) = 0\}$ is the region marked as occluded in the occlusion segmentation map S and $|Y|$ denotes the area of a region Y.

3) Output α.

7.4.6.3 Mapping the computation result to the target range of the quality component

The native component measure, α, has smaller-is-better semantics. It shall be mapped to a quality component value using:

Q = ROUND(100(1 − α))

7.4.7 Face occlusion prevention

7.4.7.1 Description

This quality measure refers to the requirements in ISO/IEC 19794-5:2011, 5.5.7 (relevant for UC2 as specified in Reference [64]) and ISO/IEC 39794-5:2019, D.1.4.3.5 (relevant for UC1 as specified in Reference [65]).

For UC1, the head covering text of ISO/IEC 39794-5:2019, D.1.4.3.5 requires the region of the face, from the crown to the base of the chin, and from ear-to-ear, to be clearly visible.

7.4.7.2 Computation of the native quality measure

The algorithm takes as input the landmarked region F output by the algorithm in 6.6 and the face occlusion segmentation S output by the algorithm in 6.9.

1) Compute $G = \{(x,y) | F(x,y) = 1\}$.

2) If $|G| = 0$, output failureToAssess and terminate the algorithm.

3) Compute $R = \{(x,y) | S(x,y) = 0\}$.

4) Compute the fraction of the face that is occluded as $\alpha = |G \cap R|/|G|$, where |Y| denotes the area of a region Y.

5) Output α.

7.4.7.3 Mapping the computation result to the target range of the quality component

The native quality measure, α, has smaller-is-better semantics. It shall be mapped to a quality component value using:

Q = ROUND(100(1 − α))

7.4.8 Inter-eye distance

7.4.8.1 Description

This quality measure refers to the requirements in ISO/IEC 19794-5:2011, 5.7.6 (relevant for UC2 as specified in Reference [64]) and ISO/IEC 39794-5:2019, D.1.4.2.4 (relevant for UC1 as specified in Reference [65]).

For UC1, ISO/IEC 39794-5:2019, D.1.4.2.4 requires the inter-eye distance in the captured photo to be at least 90 pixels for legacy applications, and, for new passport application processes, at least 240 pixels.

The inter-eye distance (IED), measured in pixels, is related to the spatial sampling rate of the imaging hardware/software system, the camera-subject distance, the size of the subject and the yaw angle.

NOTE IED does not express resolution because resolution will be reduced by defocus, blur and compression even when the spatial sampling rate is fixed. IED is useful if those factors have been addressed by careful design and configuration.

7.4.8.2 Computation of the native quality measure

1) Compute the inter-eye distance, d_{IED}, using the algorithm specified in 6.10.

2) If d_{IED} is NaN, then return failureToAssess.

3) Otherwise, return d_{IED}.

7.4.8.3 Mapping the computation result to the target range of the quality component

The native quality measure, d_{IED}, has lower-is-worse semantics. It shall be mapped to a quality component value using:

$$Q = \mathrm{ROUND}(100\ \mathrm{SIGMOID}(d_{\mathrm{IED}}, 70, 20))$$

NOTE This function progressively penalizes low IED and has been defined in order follow several constraints: First, to fit to actual face recognition performance, second to improve next generation algorithms in order to use more information like skin texture, and thirdly to take into account the presence of human evaluation in some decision use cases.

7.4.9 Head size

7.4.9.1 Description

This quality measure refers to the requirements in ISO/IEC 19794-5:2011, 5.7.6 (relevant for UC2 as specified in Reference [64]) and ISO/IEC 39794-5:2019, D.1.4.4 (relevant for UC1 as specified in Reference [65]).

For UC1, ISO/IEC 39794-5:2019, D.1.4.4 and Table D.8 regulate size and position of the face in the image, to avoid instances of cropping or faces being too close to the edge of the image.

7.4.9.2 Computation of the native quality measure

The algorithm takes as input the landmarks L_0, ..., L_{97} output by the algorithm in 6.5.

1) Compute the distance T between the midpoint between the eyes and the chin as:

$$T = \left\| \frac{L_{60} + L_{64} + L_{68} + L_{72}}{4} - L_{16} \right\|_2$$

2) Read the height of the image, B.

3) Compute and output $D = T/B$.

7.4.9.3 Mapping the computation result to the target range of the quality component

The native quality measure, D, has two-sided semantics: low values indicate poor photo composition and very high values indicate cropping. It shall be mapped to a quality component value using:

$$Q = \mathrm{ROUND}(200\ (1 - \mathrm{SIGMOID}(|D - 0.45|, 0, 0.05))$$

7.4.10 Crop of the face image

7.4.10.1 Leftward crop of face in image

7.4.10.1.1 Description

This quality measure refers to the requirements in ISO/IEC 39794-5:2019, D.1.4.4 and D.1.4.5.4 (relevant for UC1 as specified in Reference [65]).

For UC1, ISO/IEC 39794-5:2019, D.1.4.4 indicates that cropping and mis-centring of faces can lead to recognition failure.

7.4.10.1.2 Computation of the native quality measure

1) Compute the subject's right eye centre (i.e. the eye on the left from the observer perspective), (X_R, Y_R), and inter-eye distance, d_{IED}, using the procedure given in 6.10.

2) Compute and output $v = X_R / d_{IED}$.

7.4.10.1.3 Mapping the computation result to the target range of the quality component

The native quality measure, v, shall be mapped to a quality component value using:

$$Q = \text{ROUND}(100\ \text{SIGMOID}\ (v, 0.9, 0.1))$$

7.4.10.2 Rightward crop of face in image

7.4.10.2.1 Description

This quality measure refers to the requirements given in ISO/IEC 39794-5:2019, D.1.4.4 and D.1.4.5.4 (relevant for UC1 as specified in Reference [65]).

For UC1, ISO/IEC 39794-5:2019, D.1.4.4 indicates that cropping and mis-centring of faces can lead to recognition failure.

7.4.10.2.2 Computation of the native quality measure

1) Compute the subject's left eye centre, i.e. eye on the right from the observer perspective, (X_L, Y_L), and inter-eye distance, d_{IED}, using the procedure given in 6.10.

2) Find the image width, A.

3) Compute and output $v = (A - X_L) / \mathrm{d}_{IED}$.

7.4.10.2.3 Mapping the computation result to the target range of the quality component

The native quality measure, v, shall be mapped to a quality component value using:

$$Q = \text{ROUND}(100\ \text{SIGMOID}\ (v, 0.9, 0.1))$$

7.4.10.3 Margin above the face in image

7.4.10.3.1 Description

This quality measure refers to the requirements given in ISO/IEC 39794-5:2019, D.1.4.4 and D.1.4.5.4 (relevant for UC1 as specified in Reference [65]).

For UC1, ISO/IEC 39794-5:2019, D.1.4.4 indicates that cropping and mis-centring of faces can lead to recognition failure.

7.4.10.3.2 Computation of the native quality measure

The algorithm takes as input the original landmarks L_0, ..., L_{97} output by the algorithm in 6.5.

1) Compute the left eye centre (X_L, Y_L) and the right eye centre (X_R, Y_R) using the algorithm in 6.10.

2) Compute the midpoint between the eyes as:

$$(X_C, Y_C) = \frac{(X_L, Y_L) + (X_R, Y_R)}{2}.$$

3) Compute the distance T between the midpoint between the eyes and the chin as:

$$T = \|(X_C, Y_C) - L_{16}\|_2 .$$

4) Compute and output $v = Y_C / T$

7.4.10.3.3 Mapping the computation result to the target range of the quality component

The native quality measure v shall be mapped to a quality component value using:

$$Q = \text{ROUND}(100\ \text{SIGMOID}\ (v, 1.4, 0.1))$$

7.4.10.4 Margin below the face in image

7.4.10.4.1 Description

This quality measure refers to the requirements given in ISO/IEC 39794-5:2019, D.1.4.4 and D.1.4.5.4 (relevant for UC1 as specified in Reference [65]).

For UC1, ISO/IEC 39794-5:2019, D.1.4.4 indicates that cropping and mis-centring of faces can lead to recognition failure.

7.4.10.4.2 Computation of the native quality measure

The algorithm takes as input the original landmarks L_0, ..., L_{97} output by the algorithm in 6.5.

1) Compute the left eye centre (X_L, Y_L) and the right eye centre (X_R, Y_R) using the algorithm in 6.10.

2) Compute the midpoint between the eyes as:

$$(X_C, Y_C) = \frac{(X_L, Y_L) + (X_R, Y_R)}{2}$$

3) Compute the distance T between the midpoint between the eyes and the chin as:

$$T = \|(X_C, Y_C) - L_{16}\|_2$$

4) Find the image height, B.

5) Compute and output $v = (B - Y_C) / T$

7.4.10.4.3 Mapping the computation result to the target range of the quality component

The native quality measure v shall be mapped to a quality component value using:

Q = ROUND(100 SIGMOID (v, 1.8, 0.1))

7.4.11 Head pose

7.4.11.1 General

Non-frontal head orientation will generally degrade recognition in an absolute sense because large angles can undermine face detection, and in a relative sense because recognition against reference photos is usually performed against a frontal photo.

Pose is the orientation of the head relative to the optical axis. It is expressed, as in ISO/IEC 39794-5:2019, by three rotation angles, yaw, pitch and roll about axes passing through the biometric capture subject's head. Absolute pose in an image is influential on accuracy.

The following subclauses therefore establish quality component computations that express the extent to which estimated pose (following the algorithm defined in 6.11) aligns with the canonical frontal view.

7.4.11.2 Head pose angle yaw frontal alignment

7.4.11.2.1 Description

This quality measure refers to the requirements in ISO/IEC 19794-5:2011, 5.5.9.1 and 7.2.2, (relevant for UC2 as specified in Reference [64]) and ISO/IEC 39794-5:2019, D.1.4.3.1 and D.1.4.5.3, (relevant for UC1 as specified in Reference [65]).

For UC1, ISO/IEC 39794-5:2019, D.1.4.3.1 requires that the yaw angle of the head is less than ±5° from frontal.

Yaw estimation generally proceeds on images exhibiting compound rotation in which pitch and roll can also be non-zero.

7.4.11.2.2 Computation of the native quality measure

The quality component shall be computed as described in 6.11.

NOTE 1 The angles exist in a coordinate system given in ISO/IEC 39794-5:2019, 7.21.

NOTE 2 Acceptable values are not defined in this document. For UC1 in adults, see subclause in ISO/IEC 39794-5:2019, D.1.4.3.1 and, for children "six years and under", see ISO/IEC 39794-5:2019, D.1.4.5.3.

7.4.11.2.3 Mapping the computation result to the target range of the quality component

Small absolute values of the native quality measure, i.e. the yaw angle, ϕ_{yaw} are associated with improved recognition against canonical images. The native quality measure shall be mapped to a quality component value using:

Q = ROUND(100 max(0, cos ϕ_{yaw})^2)

7.4.11.3 Head pose angle pitch frontal alignment

7.4.11.3.1 Description

This quality measure refers to the requirements in ISO/IEC 19794-5:2011, 5.5.9.2 and 7.2.2 (relevant for UC2 as specified in Reference [64]) and ISO/IEC 39794-5:2019, D.1.4.3.1 and D.1.4.5.3 (relevant for UC1 as specified in Reference [65]).

For UC1, ISO/IEC 39794-5:2019, D.1.4.3.1 requires the pitch of the head to be less than ±5° from frontal.

The sign of the pitch angle is defined according to ISO/IEC 39794-5:2019, 7.21. However, the definition of pitch value zero deviates from the definition in ISO/IEC 39794-5:2019, 7.21, where it was defined by the Frankfurt Horizon. The reason for this deviation is that the Frankfurt Horizon a) cannot be measured without sight of the subject's ears, and b) differs by an angle that is a random variable from what a human observer or a trained photographer would stipulate as being fully frontal when asking a subject to look at a far object. Instead of basing the definition of the pitch on the Frankfurt Horizon, the computation of this quality component aims to output a pitch angle zero for a pose typically assumed by the subject when asked to look in the horizontal direction in the most natural way.

7.4.11.3.2 Computation of the native quality measure

The quality component shall be computed as described in 6.11.

7.4.11.3.3 Mapping the computation result to the target range of the quality component

Small absolute values of the native quality measure, i.e. the pitch angle ϕ_{pitch}, are associated with improved recognition against canonical images. The native quality measure shall be mapped to a quality component value using:

$$Q = \text{ROUND}(100 \max(0, \cos \phi_{\text{pitch}})^2)$$

NOTE Acceptable values are not defined in this document. For UC1, a deployer of a QAA can see recommended angles for adults in ISO/IEC 39794-5:2019, D.1.4.3.1, and, for children "six years and under", in ISO/IEC 39794-5:2019, D.1.4.5.3.

7.4.11.4 Head pose angle roll frontal alignment

7.4.11.4.1 Description

This quality measure refers to the requirements in ISO/IEC 19794-5:2011, 5.5.9.3 and 7.2.2 (relevant for UC2 as specified in Reference [64]) and SO/IEC 39794-5:2019, D.1.4.3.1 and D.1.4.5.3 (relevant for UC1 as specified in Reference [65]).

For UC1, ISO/IEC 39794-5:2019, D.1.4.3.1 requires the roll of the head to be less than ±8° from frontal.

Head roll or in-plane rotation of the head is often conceived of as not problematic for face recognition because, for images with zero yaw angle, it can be corrected easily using simple rotation about an axis parallel to the optical axis.

7.4.11.4.2 Computation of the native quality measure

The quality component shall be computed as described in 6.11.

7.4.11.4.3 Mapping the computation result to the target range of the quality component

The native quality measure shall be mapped to a quality component value as follows. If $|\phi_{\text{roll}}| \geq 90$ then $Q = 0$ otherwise:

$$Q = \text{ROUND}(100 \max(0, \cos \phi_{\text{roll}})^2)$$

7.4.12 Expression neutrality

7.4.12.1 Description

This quality measure refers to the requirements in ISO/IEC 19794-5:2011, 5.5.8 (relevant for UC2 as specified in Reference [64]) and ISO/IEC 39794-5:2019, D.1.4.3.2 (relevant for UC1 as specified in Reference [65]).

For UC1, ISO/IEC 39794-5:2019, D.1.4.3.2 requires the face to have a neutral expression.

A non-neutral facial expression in a photo can cause a face recognition algorithm to produce a lower similarity score when comparing with an ISO/IEC 39794-5 photo of the same biometric capture subject.

7.4.12.2 Computation of the native quality measure

The implementation should include the algorithm described in this subclause.

The algorithm uses the CNN model modified from References [51] and [52]. The modified model is available from Reference [58].

The algorithm takes as input the image I in RGB colour channel order (8 bits per channel) output by the alignment algorithm in 6.7.

1) Crop I to the rectangle spanned by the corners (144, 148) and (472, 488).
2) Normalize I with mean μ = (123.7, 116.3, 103.5) and standard deviation σ = (58.4, 57.1, 57.4) using 6.19.
3) Resize I to dimension (224, 224) using bilinear interpolation.
4) Reshape I to a 4-dimensional input tensor $T^{[1]}$ of dimensions (1, 3, 224, 224).
5) Feed the tensor $T^{[1]}$ to the CNN model to obtain an output tensor $T'^{[1]}$ of dimensions (1, 1 280).
6) Cast the output tensor $T'^{[1]}$ to a feature vector F_1 of length 1 280.
7) Resize I to dimension (260, 260) using bilinear interpolation.
8) Reshape I to a 4-dimensional input tensor $T^{[2]}$ of dimensions (1, 3, 260, 260).
9) Feed the tensor $T^{[2]}$ to the CNN model to obtain an output tensor $T'^{[2]}$ of dimensions (1, 1 408).
10) Cast the output tensor $T'^{[2]}$ to a feature vector F_2 of length 1 408.
11) Concatenate F_1 and F_2 to obtain the feature vector F of length 2 688.
12) Run the AdaBoost classifier model on the feature vector F to obtain the output x.
13) Output x.

7.4.12.3 Mapping the computation result to the target range of the quality component

The native quality measure shall be mapped to a quality component value as follows:

$$Q = \mathrm{ROUND}(100 \cdot \mathrm{SIGMOID}(x, -5000, 5000))$$

7.4.13 No head covering

7.4.13.1 Description

This quality measure refers to the requirements of ISO/IEC 19794-5:2011, 5.5.7 and 7.2.10, (relevant for UC2 as specified in Reference [64]) and ISO/IEC 39794-5:2019, D.1.4.3.5, D.1.4.3.6 and D.1.4.3.7 (relevant for UC1 as specified in Reference [65]).

For UC1, ISO/IEC 39794-5:2019, D.1.4.3.5 requires an elliptically shaped region on the face to be visible.

The purpose of this quality component is to determine whether a biometric capture subject is wearing a head covering.

7.4.13.2 Computation of the native quality measure

The algorithm takes as input the face parsing segmentation map M computed by the algorithm in 6.8.

1) Crop M from the bottom by 204 pixels.

2) Count the number n of pixels in M having value 16 or 18.

3) Output $r = n/m$ where m is the number of pixels in M.

NOTE While the label 18 in the face parsing segmentation map specifies head coverings, the label 16 specifies clothing and is used in step 2 to account for hoods put over the head.

7.4.13.3 Mapping the computation result to the target range of the quality component

The native quality measure, r, shall be mapped to a quality component value Q using the following computation.

1) Let $T_0 = 0$, $T_1 = 0.95$, $w = 0.1$, and $x_0 = 0.02$

2) If $r \leq T_0$, output Q=100 and terminate the algorithm; otherwise continue with the next step.

3) If $r \geq T_1$, output Q=0 and terminate the algorithm; otherwise, continue with the next step.

4) Let $s(x) = 1/(1 + \exp((x_0 - x)/w))$

5) Let $h(x) = s(T_1) - s(x)$

6) Let $q(x) = h(x) / h(T_0)$

7) Compute and output $Q = \mathrm{ROUND}(100\ q(r))$

8 Face image quality block

8.1 Binary encoding

In binary biometric data blocks, quality data shall be encoded as specified in ISO/IEC 39794-1:2019 or ISO/IEC 19794-1:2011.

Quality measures for face images encoded in the format specified in ISO/IEC 39794-5:2019 should be placed into the quality blocks of that format. CBEFF quality fields should not be used in place of ISO/IEC 39794-5:2019 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. Multiple quality measures calculated by the same algorithm (same quality algorithm vendor identifier and same quality algorithm identifier) shall not be present in a single biometric data interchange record.

8.2 XML encoding

In XML documents, quality data shall be encoded as specified in ISO/IEC 39794-1:2019.

8.3 Organization identifiers

The reference implementation shall use the biometric organization identifier of ISO/IEC JTC 1/SC 37, which is 257 (0101_{Hex}).

Any other implementation shall use a different organization identifier.

8.4 Algorithm identifiers

Table 7 lists the QAA identifiers for the quality measures defined in this document.

These identifiers have been registered with the Biometric Registration Authority in accordance with ISO/IEC 19785-2.[11]

A conformant QAA shall use the QAA identifiers given in Table 7.

Table 7 — QAA identifiers

Quality assessment algorithm identifier		Description	Governing subclause
Hexadecimal	**Decimal**		
41_{Hex}	65	Unified face image quality score calculated with version 1.0.0 of SC 37 approved reference implementation	7.2
42_{Hex}	66	Background uniformity	7.3.2
43_{Hex}	67	Illumination uniformity	7.3.3
44_{Hex}	68	Luminance mean	7.3.4.2
45_{Hex}	69	Luminance variance	7.3.4.3
46_{Hex}	70	Under-exposure prevention	7.3.5
47_{Hex}	71	Over-exposure prevention	7.3.6
48_{Hex}	72	Dynamic range	7.3.7
49_{Hex}	73	Sharpness	7.3.8
$4a_{Hex}$	74	No compression artefacts	7.3.9
$4b_{Hex}$	75	Natural colour	7.3.10
$4c_{Hex}$	76	Single face present	7.4.2
$4d_{Hex}$	77	Eyes open	7.4.3
$4e_{Hex}$	78	Mouth closed	7.4.4
$4f_{Hex}$	79	Eyes visible	7.4.5
50_{Hex}	80	Mouth occlusion prevention	7.4.6
51_{Hex}	81	Face occlusion prevention	7.4.7
52_{Hex}	82	Inter-eye distance	7.4.8
53_{Hex}	83	Head size	7.4.9
54_{Hex}	84	Leftward crop of face in image	7.4.10.1
55_{Hex}	85	Rightward crop of face in image	7.4.10.2
56_{Hex}	86	Margin above face in image	7.4.10.3
57_{Hex}	87	Margin below face in image	7.4.10.4
58_{Hex}	88	Head pose angle yaw frontal alignment	7.4.11.2
59_{Hex}	89	Head pose angle pitch frontal alignment	7.4.11.3
$5a_{Hex}$	90	Head pose angle roll frontal alignment	7.4.11.4
$5b_{Hex}$	91	Expression neutrality	7.4.12
$5c_{Hex}$	92	No head covering	7.4.13
$5d_{Hex}$	93	Shoulder presentation	D.3.2
$5e_{Hex}$	94	Camera to subject distance	D.3.3

Table 7 *(continued)*

Quality assessment algorithm identifier		Description	Governing subclause
Hexadecimal	**Decimal**		
$5f_{Hex}$	95	Reserved for unified face image quality scores calculated with future versions of SC 37 approved reference implementation	
60_{Hex}	96		
61_{Hex}	97		
62_{Hex}	98		
63_{Hex}	99		
64_{Hex}	100		
65_{Hex}	101		
66_{Hex}	102		
67_{Hex}	103		
68_{Hex}	104		
69_{Hex}	105		

Annex A
(normative)

Conformance test assertions

A.1 Overview

In order to achieve the objective of this document, it is necessary to test biometric products to ensure their conformance. Conformant implementations are a prerequisite for achieving interoperability among implementations. Therefore, there is a need for a standardized conformance testing methodology, test assertions, and test procedures as applicable to specific modalities addressed by this document. The test assertions cover all normative quality components, so that the conformity results produced by the test suites will reflect the real degree of conformity of the implementations to ISO/IEC 29794-5 face image quality blocks.

This annex specifies conformance test criteria, and a conformance test set to determine whether an implementation reproduces the target quality measure of the reference implementation OFIQ.

The conformance test set consists of face images collected by ISO/IEC JTC 1/SC 37, Biometrics, for the purpose of testing conformance with this document. It can be obtained from Reference [57].

NOTE For the conformance test set, Tables A.1 and A.2 show the values for the native quality measures. These values are available in comma-separated-value files.[57]

OFIQ, version 1.0.0, is a reference implementation for this document; its source code is available from Reference [59]. The pre-compiled release of this reference implementation produces no differences from the target values in Table A.3 and A.4.

A.2 Conformance test

To conform, a quality assessment implementation claiming conformance to this document shall

— run on all images in the conformance test set,

— return quality component values for all components marked as mandatory (M) in Table 3 for all use cases for which conformance is claimed, and

— produce quality component values that have absolute difference to the target quality measure values in Table A.3 and Table A.4 less than or equal to 1 (i.e. 1 % of the maximum value 100). The differences shall also be verified by independent testing institutions on testing data which is sequestered, meaning a conformance testing certificate can be issued with a certification report. The testing API should allow a variety of implementations (e.g. C++, Python).

An implementation that does not reproduce the target quality measures given in Tables A.3 and A.4 is not conformant to this document.

Table A.1 — Native quality measure values for conformance test set – capture-related measures.

Identifier	Unified quality score	Background uniformity	Illumination uniformity	Luminance mean	Luminance variance	Under-exposure prevention	Over-exposure prevention	Dynamic range	Sharpness	No compression artifacts	Natural colour
b-01-smile	24.329985	71.77712	0.141003	0.296285	0.014954	0.049163	0	6.910499	44	0.947065	0
b-02-rolled	24.698753	66.769635	0.166955	0.298599	0.014252	0.055943	0	6.888223	50	0.944888	0.171868
b-03-headcovered	16.142765	41.745187	0.117539	0.322942	0.04005	0.10221	0	7.438106	50	0.96607	0
b-04-headcovered	18.811047	66.071552	0.098776	0.30336	0.019721	0.027474	0	6.880414	47	0.978856	0.516643
b-05-scarf	17.191828	82.221345	0.49827	0.298608	0.0145	0.023528	0	6.932558	45	0.886013	3.350172
b-06-mask	14.262262	38.975958	0.018166	0.497561	0.059659	0.131261	0	7.609802	50	0.938479	4.805962
b-07-glasses	21.134285	66.56598	0.106939	0.318891	0.014461	0.017577	0	6.869207	38	0.960998	0.948361
b-08-sunglasses	18.325424	79.097756	0.149691	0.2861	0.024641	0.044249	0	6.993652	47	0.944594	0
b-09-background	24.752863	242.78947	0.030277	0.250153	0.028186	0.173227	0	7.025578	47	0.953989	0
c-01-frontal	26.449306	61.625131	0.320816	0.282818	0.00732	0.0257	0	6.451986	49	0.955245	0
c-02-headcovered-mouthopen	14.149646	52.290763	0.842057	0.316663	0.047588	0.13586	0	7.417117	48	0.978499	0
c-03-headcovered	13.33449	49.233287	0.838384	0.340252	0.054218	0.141956	0	7.504903	48	0.964031	0
c-04-headcovered	17.668522	61.194849	0.283747	0.277023	0.01725	0.013022	0	6.566232	45	0.973624	0
c-05-scarf	17.155811	69.987374	0.385813	0.189517	0.010255	0.070675	0	6.549825	44	0.902307	100
c-06-glasses	23.296429	68.311044	0.202938	0.248062	0.012939	0.0423	0	6.849478	46	0.972348	0
c-07-twofaces	26.875198	87.475872	0.274221	0.295112	0.015774	0.037523	0	6.981078	47	0.956223	0
c-08-sunglasses	22.899508	43.20658	0.267747	0.271406	0.021515	0.021937	0	6.847849	40	0.961401	0
c-09-mask	18.188141	35.980417	0.012111	0.443931	0.067912	0.241559	0	7.459769	48	0.954354	4.726742
c-10-background	27.733389	232.17879	0.280277	0.240043	0.021413	0.118626	0	6.931612	49	0.974399	0
r-01-frontal	26.390884	48.120312	0.196759	0.190488	0.009668	0.105363	0	6.433408	49	0.961691	0
r-02-rolled	25.76358	92.862484	0.207347	0.190441	0.008405	0.082671	0	6.314165	48	0.965204	0
r-03-mouthopen	26.416	48.339777	0.19213	0.18692	0.00863	0.094275	0	6.252619	45	0.973072	0
r-04-headcovered-eyeclosed	20.354622	46.267522	0.20216	0.225396	0.007762	0.032277	0	6.338847	48	0.965957	0
r-05-headcovered	18.509974	41.910229	0.208163	0.198766	0.022266	0.239812	0	6.712063	49	0.968723	0
r-06-pitch	25.199898	55.484277	0.229167	0.177005	0.00872	0.103467	0	6.209496	44	0.968835	0
r-07-scarf	23.497959	44.647808	0.114969	0.258515	0.013047	0.024735	0	6.729036	48	0.976354	0
r-08-mask	16.549667	67.653893	0.388504	0.103719	0.022853	0.087205	0	5.55899	41	0.943364	100
r-09-background	27.223253	187.4076	0.020062	0.225448	0.030372	0.351883	0.000065	7.006914	47	0.97056	0

Table A.2 — Native quality measure values for conformance test set – subject-related measures

Identifier	Single face present	Eyes open	Mouth closed	Eyes visible	Mouth occlusion prevention	Face occlusion prevention	Inter-eye distance	Head-size	Leftward crop of the face image	Rightward crop of the face image	Margin above of the face image	Margin below of the face image	Head pose yaw angle	Head pose pitch angle	Head pose roll angle	Expression neutrality	No head covering
b-01-smile	0	0,07022	0,002466	0	0	0,004408	493,556527	0,224851	3,428939	3,698472	2,426863	2,022937	-1,20676	8,573003	1,415725	-10990,88379	0
b-02-rolled	0	0,074475	0,002642	0	0	0,0049	446,656552	0,209942	3,610465	4,116961	2,401876	2,364051	-2,582642	1,180704	16,291553	10100,62988	0
b-03-headcovered	0	0,060311	0,001891	0,546753	0	0,192342	388,968275	0,185504	4,02158	4,768703	2,491129	2,903195	-3,703697	6,903509	-0,161357	660,10614	0,401441
b-04-headcovered	0	0,082626	0,001407	0,15439	0	0,136651	403,257111	0,176257	3,882407	4,619196	2,520487	3,156945	-1,596962	4,224815	1,186767	-5549,105469	0,507551
b-05-scarf	0	0,089949	0,065274	0	1	0,52303	431,681985	0,18622	3,79167	4,280541	1,915176	3,458512	-7,966364	7,258156	1,776166	-1636,364746	0
b-06-mask	0	0,075264	0,027464	0,015196	1	0,73643	482,834934	0,225763	3,471587	3,797178	1,741202	2,688153	-2,938281	1,624212	2,442354	-4067,674561	0
b-07-glasses	0	0,082377	0,001374	0,380124	0	0,131707	440,181512	0,180528	3,734014	4,134005	2,809487	2,729805	-1,603732	13,847975	4,400784	8202,913086	0
b-08-sunglasses	0	0,065409	0,000997	1	0	0,200506	558,847682	0,248645	3,11841	3,299421	1,548853	2,474971	-3,199275	-2,531154	1,624107	8907,189453	0
b-09-background	0	0,089511	0,00416	0	0	0,003397	519,766945	0,238479	3,543728	3,276168	2,470235	1,725211	-1,803813	13,81824	1,38912	-241,303101	0
c-01-frontal	0	0,06652	0,003328	0	0	0,003762	375,044452	0,166626	4,546408	4,517076	2,200024	3,801665	0,636195	5,126308	1,723877	-2553,50293	0
c-02-headcovered-mouthopen	0	0,059627	0,082712	0,999656	0	0,276124	435,855168	0,210091	3,604512	4,347023	1,768317	2,994444	-3,567957	14,853941	3,603158	-6863,739746	0,462296
c-03-headcovered	0	0,061926	0,004447	1	0	0,326122	429,707512	0,201075	3,721792	4,325389	1,979702	2,996707	-3,061079	11,878896	1,250443	-7114,487793	0,411454
c-04-headcovered	0	0,067124	0,049825	0,482405	0	0,159333	449,020387	0,204089	3,68705	4,062004	1,928526	2,974259	-3,75841	12,207628	6,754395	-12166,3457	0,515918
c-05-scarf	0	0,08695	0,040533	0,000881	1	0,634053	609,265618	0,215554	2,560953	3,411862	2,269089	2,370346	-4,255126	16,617661	6,058564	-1136,262085	0
c-06-glasses	0	0,050625	0,002157	0,450213	0	0,10081	590,141115	0,257166	2,667819	3,463753	2,140145	1,750326	-3,157394	8,111413	1,426863	-1481,442627	0
c-07-twofaces	0,333736	0,067793	0,002632	0	0	0,006484	472,73993	0,210711	3,530623	3,871613	2,230722	2,518101	-2,851403	10,251065	4,484339	-7237,871582	0
c-08-sunglasses	0	0,067954	0,005753	1	0	0,182349	408,084326	0,172441	4,27143	4,139097	3,208191	2,595051	-0,625034	9,825079	3,159786	1323,649902	0
c-09-mask	0	0,05741	0,009154	0,076569	1	0,707127	443,230718	0,197248	3,960616	3,863576	2,562638	2,507311	1,321376	11,545669	5,036475	-3031,031494	0
c-10-background	0	0,081871	0,001497	0	0	0,001576	374,239097	0,165638	4,613603	4,47972	2,307622	3,734152	-3,695724	4,17495	4,768653	5935,286133	0
r-01-frontal	0	0,078848	0,006977	0	0	0,087116	537,077622	0,224826	3,093062	3,53812	1,818983	2,631345	-0,920045	-0,251555	-1,3356	6674,853516	0
r-02-rolled	0	0,060196	0,002228	0	0	0,084338	488,587618	0,222656	3,709694	3,381765	2,033863	2,459582	-3,01603	1,360564	-27,373343	4869,32373	0
r-03-mouthopen	0	0,066931	0,08053	0	0	0,083542	466,756161	0,215607	3,660207	3,845261	1,918722	2,71934	-5,343358	-0,621376	-4,113008	-15135,90137	0
r-04-headcovered-eyeclosed	0	0,023177	0,005358	0	0	0,086418	454,949992	0,207003	3,705352	3,963403	2,392823	2,440775	-4,733527	3,48683	-3,996515	-11079,18359	0,417207
r-05-headcovered	0	0,05979	0,002813	0,189971	0	0,158896	428,228233	0,197116	4,205831	3,891333	2,306325	2,766835	-5,747996	1,303665	-0,029483	-7027,903809	0,301582
r-06-pitch	0,316289	0,086448	0,003407	0,193437	0	0,102952	566,373838	0,230215	3,242948	3,106719	2,116965	2,226854	-3,647577	10,990835	1,270625	-7541,475098	0
r-07-scarf	0	0,085567	0,001838	0	0	0,197904	401,527105	0,134956	4,209005	4,331186	2,740027	4,669623	-2,808316	0,74684	-1,608323	6770,018066	0
r-08-mask	0	0,111367	0,019661	0,021734	1	0,72939	597,589439	0,202228	2,922283	3,136517	1,983116	2,961797	-1,10636	1,369975	3,862225	-2507,209473	0
r-09-background	0	0,09236	0,003772	0,079576	0	0,096862	353,583928	0,147008	5,153209	4,398046	3,551485	3,251169	0,575067	5,635712	1,961695	-924,118652	0

Table A.3 — Quality component values for conformance test set – capture-related measures.

Identifier	Unified quality score	Background uniformity	Illumination uniformity	Luminance mean	Luminance variance	Under-exposure prevention	Over-exposure prevention	Dynamic range	Sharpness	No compression artefacts	Natural colour
b-01-smile	63	67	56	87	100	100	100	86	99	100	100
b-02-rolled	66	69	58	88	99	100	100	86	100	100	99
b-03-headcovered	7	80	53	92	80	100	100	93	100	100	100
b-04-headcovered	17	69	50	89	99	100	100	86	100	100	97
b-05-scarf	10	62	81	88	99	100	100	87	100	100	83
b-06-mask	3	81	30	100	63	100	100	95	100	100	76
b-07-glasses	33	69	51	92	99	100	100	86	99	100	95
b-08-sunglasses	14	63	57	85	95	100	100	87	100	100	100
b-09-background	66	17	35	73	92	100	100	88	100	100	100
c-01-frontal	79	71	71	84	82	100	100	81	100	100	100
c-02-headcovered-mouthopen	3	75	95	91	73	100	100	93	100	100	100
c-03-headcovered	2	77	95	94	67	100	100	94	100	100	100
c-04-headcovered	11	71	69	82	100	100	100	82	100	100	100
c-05-scarf	10	67	75	45	93	100	100	82	99	100	0
c-06-glasses	53	68	62	72	98	100	100	86	100	100	100
c-07-twofaces	82	60	68	87	100	100	100	87	100	100	100
c-08-sunglasses	49	79	67	81	98	100	100	86	99	100	100
c-09-mask	14	83	27	99	58	100	100	93	100	100	77
c-10-background	86	19	68	69	98	100	100	87	100	100	100
r-01-frontal	79	77	61	45	91	100	100	80	100	100	100
r-02-rolled	74	58	62	45	87	100	100	79	100	100	100
r-03-mouthopen	79	77	61	43	88	100	100	78	100	100	100
r-04-headcovered-eyeclosed	27	78	62	62	84	100	100	79	100	100	100
r-05-headcovered	15	80	62	49	97	100	100	84	100	100	100
r-06-pitch	70	74	64	39	88	100	100	78	99	100	100
r-07-scarf	55	79	52	76	98	100	100	84	100	100	100
r-08-mask	8	68	75	13	97	100	100	69	99	100	0
r-09-background	84	28	31	62	90	100	99	88	100	100	100

Table A.4 — Quality component values for conformance test set – subject-related measures.

Identifier	Single face present	Eyes open	Mouth closed	Eyes visible	Mouth occlusion prevention	Face occlusion prevention	Inter-eye distance	Head-size	Leftward crop of the face image	Rightward crop of the face image	Margin above of the face image	Margin below of the face image	Head pose yaw angle	Head pose pitch angle	Head pose roll angle	Expression neutrality	No head covering
b-01-smile	100	99	96	100	100	100	100	2	100	100	100	90	100	98	100	23	100
b-02-rolled	100	100	96	100	100	100	100	2	100	100	100	100	100	100	92	95	100
b-03-headcovered	100	98	96	45	100	81	100	1	100	100	100	100	100	99	100	76	4
b-04-headcovered	100	100	96	85	100	86	100	1	100	100	100	100	100	99	100	47	1
b-05-scarf	100	100	90	100	0	48	100	1	100	100	99	100	98	98	100	66	100
b-06-mask	100	100	95	98	0	26	100	2	100	100	97	100	100	100	100	55	100
b-07-glasses	100	100	96	62	100	87	100	1	100	100	100	100	100	94	99	93	100
b-08-sunglasses	100	99	96	0	100	80	100	4	100	100	82	100	100	100	100	94	100
b-09-background	100	100	96	100	100	100	100	3	100	100	100	32	100	94	100	72	100
c-01-frontal	100	99	96	100	100	100	100	1	100	100	100	100	100	99	100	62	100
c-02-headcovered-mouthopen	100	98	88	0	100	72	100	2	100	100	98	100	100	93	100	41	2
c-03-headcovered	100	99	96	0	100	67	100	1	100	100	100	100	100	96	100	40	4
c-04-headcovered	100	99	92	52	100	84	100	1	100	100	99	100	100	96	99	19	1
c-05-scarf	100	100	93	100	0	37	100	2	100	100	100	100	99	92	99	68	100
c-06-glasses	100	96	96	55	100	90	100	4	100	100	100	38	100	98	100	67	100
c-07-twofaces	67	99	96	100	100	99	100	2	100	100	100	100	100	97	99	39	100
c-08-sunglasses	100	99	96	0	100	82	100	1	100	100	100	100	100	97	100	78	100
c-09-mask	100	98	96	92	0	29	100	1	100	100	100	100	100	96	99	60	100
c-10-background	100	100	96	100	100	100	100	1	100	100	100	100	100	99	99	90	100
r-01-frontal	100	100	96	100	100	91	100	2	100	100	99	100	100	100	100	91	100
r-02-rolled	100	98	96	100	100	92	100	2	100	100	100	100	100	100	79	88	100
r-03-mouthopen	100	99	88	100	100	92	100	2	100	100	99	100	99	100	99	12	100
r-04-headcovered-eyeclosed	100	58	96	100	100	91	100	2	100	100	100	100	99	100	100	23	3
r-05-headcovered	100	98	96	81	100	84	100	1	100	100	100	100	99	100	100	40	10
r-06-pitch	68	100	96	81	100	90	100	2	100	100	100	99	100	96	100	38	100

Table A.4 *(continued)*

Identifier	Single face present	Eyes open	Mouth closed	Eyes visible	Mouth occlusion prevention	Face occlusion prevention	Inter-eye distance	Head-size	Leftward crop of the face image	Rightward crop of the face image	Margin above of the face image	Margin below of the face image	Head pose yaw angle	Head pose pitch angle	Head pose roll angle	Expression neutrality	No head covering
r-07-scarf	100	100	96	100	100	80	100	0	100	100	100	100	100	100	100	91	100
r-08-mask	100	100	95	98	0	27	100	1	100	100	100	100	100	100	100	62	100
r-09-background	100	100	96	92	100	90	100	0	100	100	100	100	100	99	100	69	100

Annex B
(informative)

Quantitative goal for face image QAAs

B.1 Introduction

ISO/IEC 29794-1 delineates three aspects of the umbrella term "quality":

— Character: This is a statement of the distinctiveness of the anatomical biometric characteristic, e.g. a scarred fingerprint or a heavily bearded face can have poor character.

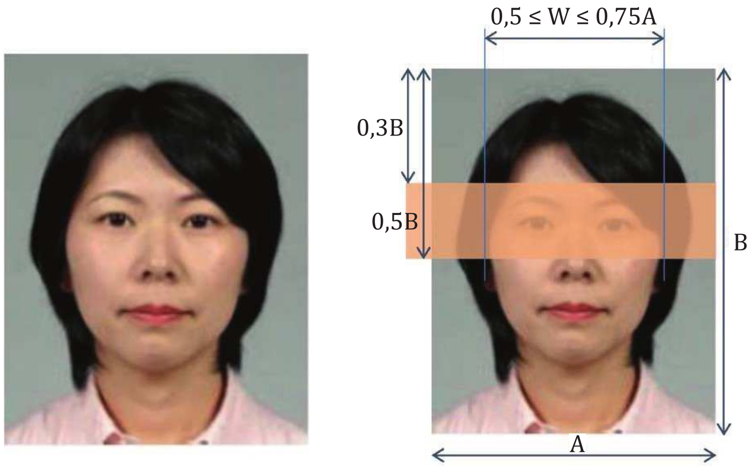
— Fidelity: This is any measurement that indicates how well a captured digital image faithfully represents the analogue source, e.g. a blurred image of a face omits detail and has low fidelity.

— Utility: This is used to indicate the value of an image to a receiving recognition algorithm. This aspect is the most relevant in this document.

This document conceives of quality scores as being measures of utility rather than, for example, fidelity, because utility of a sample to a recognition engine is what motivates end-users and because it gives a quantitative goal for development of quality scores, in the supervised machine learning sense. The efficacy of this approach for use with fingerprint systems has been demonstrated with the NIST Fingerprint Image Quality Algorithm.[10]

NOTE This approach has also been adopted for fingerprints. ISO/IEC 29794-4 defines the NFIQ algorithm,[10] which was trained using a machine learning scheme to be a predictor of fingerprint recognition comparison accuracy. That algorithm, and its commercial analogues, have been run tens of billions of times in large scale identity operations in many global programs, including Aadhaar (India) and immigration (USA).

However, recognition outcomes depend on the property of at least two images, not just the sample being submitted to a QAA. This apparent disconnect is handled by the following assumption: an image quality assessment algorithm can be designed to return quality values that reflect the expected comparison outcome of the target image with a canonical high-quality portrait image of the form given in Figure B.1.



a) Example image **b) Face geometry**

Figure B.1 — Canonical portrait photograph (ISO/IEC 39794-5)

Formally, if a face verification algorithm, V, compares two samples X_1 and X_2, to produce a comparison score:

$$S = \mathrm{V}(X_1, X_2)$$

this document conceives of QAAs predicting S from X_1 alone but under the assumption that X_2 would be a canonical portrait image, i.e. a pristine image of the same subject that is fully conformant to ISO and ICAO specifications.

NOTE 1 Although the target in Figure B.1 is a frontal photograph, quality assessment can also be performed referencing a different standard view of a face. In certain scenarios, a different view can in fact be desirable, for example, forensic face examiners have preferred views where the ear is visible (it is for this reason that the immigration agencies in the United States used to require a quarter-left view on identity documents). However, in the present example, the target is the ISO/ICAO portrait because the face recognition industry is currently capitalized on the basis of frontal face recognition. This document could potentially be extended to adopt quality assessment against an alternative standardized view.

Thus, a QAA F operating on an image X_1 produces value Q = F(X_1), which in the statistical sense predicts S because it implicitly assumes the comparison V(X_1, X_{PORTRAIT}). Quality scores are evaluated as predictors of mated comparison scores; this is difficult because they are "blind" to the hidden portrait image in the comparison. This goal respects the ISO/IEC 39794-5 specification as the reference standard for automated face recognition.

NOTE 2 The term "blind" is borrowed from the image fidelity literature in which a "blind PSNR" i.e. peak signal to noise ratio is computed from, for example, a JPEG image or a video clip as a statement of quality. Such techniques can have applicability here.

Without this formulation of the quality problem, the position, noted in the academic literature, that quality assessment cannot be performed on a single image (i.e. that quality should "come in pairs") would be correct. Such assertions note that recognition outcomes (that are the result of comparing a pair of images) depend on the properties of both images. For example, when face recognition algorithms are used to compare images with different head pose yaw angles, three behaviours occur: some algorithms give high FNMR except when the two images are frontal. Other algorithms give low FNMR for images with approximately the same yaw angle, even if non-frontal. Finally, today, some algorithms exist that offer considerable pose invariance. The purpose of this discussion is to demonstrate that recognition outcome can depend on the pair of images, but quality assessment, run on a single image (potentially long before any recognition occurs) implicitly assumes the reference photo will be an ISO/ICAO portrait.

B.2 Quality value as predictor of recognition performance

Quality values that predict recognition outcomes can be evaluated by plotting accuracy measures for groups of samples of particular quality. This may be achieved using an error trade-off characteristic, FNMR vs. FMR parametric on threshold, or by plotting both FNMR vs. threshold and FMR vs. threshold. Such plots will reveal changes in error rates as low-quality images are discarded from both mated and non-mated comparisons.

NOTE Experimental and analytic results have long shown that mated and non-mated score distributions are not independent. Sample quality affects both distributions, although not necessarily equally.

B.3 Algorithm dependence

This document specifies methods that, if implemented and used:

— are intended to reduce false negative errors below the number that would be present if the methods had not been used, noting that false negative errors can also occur due to changes in images that are not quality-related, e.g. long-term ageing;

— can reduce false positive errors, noting that false positives can occur even in very high-quality images, e.g. from twins or other demographic groups.

This document advances quality measures that in many cases quantify conformance to the ISO/IEC 39794-5 specifications for a frontal image. The reason for supporting conformance is to improve the likelihood that two images of a person conform to that standard, thereby looking more similar, and thereby improving similarity scores and reducing false negative outcomes.

Recognition algorithms extract various proprietary features from face images and have different accuracies and tolerance of quality problems. However, given extreme degradations they all fail: Sufficiently over- or under-exposed images will cause false negatives; blurred faces, likewise; faces presented at high pitch or yaw angles will generally cause failure. The approach in building a QAA, and in testing it, is to predict failure from a suite of recognition algorithms.

Annex C
(informative)

Applications of quality measures

C.1 Overview

C.1.1 General

This Annex describes four primary use cases for scalar quality scores and quality components.

C.1.2 Photo acceptance

Quality measures can be used to discard or retain an image for further use in a biometric recognition process, either directly against a given threshold or, in the case of quality vectors, using a given function. If a particular aspect of an image's quality is too low, a system can potentially discard the image and initiate collection of a new image

EXAMPLE 1 The acquisition could be completed as soon as the unified quality score from subclause 7.2 is above a given threshold.

EXAMPLE 2 Some subject-related metrics (pose, mouth open) can be computed to provide on-the-fly feedback to the capture subject.

Such processing can be implemented in a camera, in a client computer, or on a remote server. Such a capability is most useful during initial enrolment, when a prior reference image of the biometric capture subject is not available. It is also useful when forwarding the image to a remote recognition service would be time consuming or expensive.

The photo acceptance methodology will vary depending on the type of use-case:

— ID credential enrolment — In this use-case, the face image will be printed on a document, used for a maximum of 10 years and should support human examination. Image quality should be as high as possible. All quality measures of Table 3 should be computed and should exceed some minimum thresholds corresponding to what is required in ISO/IEC 39794-5:2019, Clause D.1. Whether the photo is ultimately accepted or discarded on the basis of a low quality score may be confirmed by a human operator later in the process.

— One-time usage probe sample with instantaneous response — In this use-case, the quality should be high enough to meet the intended level of accuracy of the system while optimizing the capture subject convenience.

  — For example, if a device produces blurred images, it should be fixed; if lightning environment is not good, it should be improved. This is strong support for the use of reified quality measures, as opposed to use of those generated from abstract AI mechanisms.

  — Quality component values should not have individual thresholds. Each QAA would make quality assessment errors, and even with low discarding errors for each, these errors would accumulate. In that case, as current face recognition systems have a very high accuracy, the failure to acquire due to incorrect quality assessment could easily exceed the false non-match rate, effectively degrading the system performance.

— Other use-cases — Enrolment for other usage, delayed verification, probe later used as enrolment:

  — In these use-cases, quality should be high enough to ensure later usage and interoperability, but the convenience trade off will depend on the exact use-case.

— An efficient methodology could be to combine the global quality with a selection of the most relevant quality component values, which should be verified against certain thresholds.

— As a global rule, to ensure good user experience, an acquisition should not be discarded by an acquisition-related metric, which the capture subject cannot correct, as long as the global quality score is above a defined threshold.

— Subject-related measures may be computed to provide feedback on-the-fly. Additionally, a selection of subject-related measures which have a direct impact on accuracy should be checked against certain minimum thresholds. This could include for example horizontal and vertical position, inter-eye distance, mouth-closed, eye visibility and pose quality components.

— As current face recognition algorithms are resilient to most quality defects, and to ensure good capture subject experience, these thresholds should be fixed so that the rate at which portraits are incorrectly discarded for a given quality component remain low. Otherwise, quality assessment errors would accumulate and the total failure to acquire rate of the system would be high, with only limited improvement of the recognition accuracy.

  — For example, the portrait of a capture subject with their mouth closed should not be incorrectly discarded in more than 0.1 % of cases when checking the mouth-closed quality component from 7.4.4.

  — As another example, pose should be sufficiently frontal. Current face recognition algorithms can provide good pose invariance, so thresholds should be set according to the use-case.

  — The photo should be finally accepted if all the selected subject-related quality components and the global quality score exceed their respective thresholds.

C.1.3 Monitoring of the imaging process

Acquisition-related metrics, which the capture subject cannot correct, may be computed and used for monitoring.

C.1.4 Quality summarization

Quality values may be useful as a management indicator. For example, in an enterprise where face images are being collected from many biometric capture subjects (e.g. by different staff, at different sites, under different conditions), the unified quality score and quality components can be aggregated to summarize the effectiveness of the collection. This can be achieved using a statistic such as the mean, or the proportion with a low or high value. Such summarization can be used to reveal site-specific problems or population effects, or as a response variable in A-B tests or to reveal trends, diurnal or seasonal variation.

NOTE In cases where samples are collected from the same persons regularly(for example, in a frequent traveller system) aggregated results from the comparison of mated image pairs can be an excellent indicator of expected recognition performance and can reveal image quality variations across time, collection sites, etc.

C.1.5 Photo selection

Given $K > 1$ images of a person, values can be used in the selection of the best image. This is useful when a capture process includes some variation, e.g. due to unavoidable motion of the subject or camera. This operation is useful when a receiving system expects exactly one image. The capture subsystem can institute a minimum quality threshold starting a new acquisition if the threshold is not met, or it may simply determine which of the several collected images is of the highest quality.

NOTE Ordinarily, this function does not replace recognition. Thus, in an identification application, a system would generally enrol all K images of a person rather than select just one. This avoids the possibility of the quality assessment infrastructure being an imperfect predictor of recognition outcome, because it can arise that an enrolled image with lower quality produces a high similarity score when compared to probe images that have idiosyncratic characteristics, such as extreme view angle or exaggerated facial expression. However, if some images were collected decades ago, then ageing can reduce the utility of the image to a recognition against a recent image, even if quality is excellent.

C.1.6 Weighting partial results in multi-biometric systems

Quality values can be used to weight samples in repeated- or multiple-instance systems, or samples in multimodal systems.

Annex D
(informative)

Quality requirements with no quality measure

D.1 Purpose

This annex provides quality requirements from ISO/IEC 19794-5:2011 and ISO/IEC 39794-5:2019, which, at the time of publication of this document (ISO/IEC 29794-5) cannot be formulated or tested as a quality measure. It is possible that future editions of this document will contain such measures.

D.2 Imaging-related quality components

D.2.1 Radial distortion prevention

ISO/IEC 19794-5:2011, 7.3.6 requires the “fish eye” effect associated with wide angle lenses to be absent. Further, ISO/IEC 39794-5:2019, D.1.4.2.2 requires the maximum magnification distortion rate to be at or below 7 %.

D.2.2 Pixel aspect ratio

ISO/IEC 19794-5:2011, 7.4.1.1 requires digital cameras and scanners to produce images with a pixel aspect ratio of 1:1. Equivalently, ISO/IEC 39794-5:2019, D.1.4.2.2 requires cameras for which the vertical and horizontal pixel density is the same. Small deviations from these specifications are thought to have little impact on recognition.

D.3 Subject-related quality components

D.3.1 Frontal gaze

D.3.1.1 Description

The capture subject should be looking directly at the camera.

D.3.1.2 Computation of the native quality measure:

1) Detect a face as specified in 6.4.

2) Compute facial landmarks as specified in 6.5.

3) Let the coordinates of the pupil of the subject’s left eye, (X_L, Y_L), be the coordinates of landmark L_{97}.

4) Let the coordinates of the pupil of the subject’s right eye, (X_R, Y_R), be the coordinates of landmark L_{96}.

5) Compute the distance W_L between the inner and outer canthi of the left eye.

6) Compute the distance W_R between the inner and outer canthi of the right eye.

7) Compute the distance D_L between the inner canthus and the centre of the left eye.

8) Compute the distance D_R between the inner canthus and the centre of the right eye.

9) Compute the ratio $R_L = D_L / W_L$ for the left eye.

10) Compute the ratio $R_R = D_R / W_R$ for the right eye.

11) Map the values of R_L and R_R to a quality component value in the range [0,1], such that:

   a) R_L and R_R values in the range [0.43, 0.57] map to values close to 1.

   b) R_L and R_R values close to 0 or 1 map to values close to 0.

D.3.2 Shoulder presentation

D.3.2.1 Description

ISO/IEC 19794-5:2011, 7.2.5 requires shoulders to be “square on” to the camera, with the implication that this will be measured and corrected. The orientation of the face and shoulders can differ.

D.3.2.2 Computation of the native quality measure

An implementation shall estimate the smaller angle between a line parallel to the optical axis that intersects a line parallel to the biometric capture subject’s clavicle, φ.

D.3.2.3 Mapping the computation result to the target range of the quality component

Conformant presentation corresponds to angle φ = 90. The native quality measure shall be mapped to quality component value using:

$Q = \max(0, \mathrm{ROUND}(100 \sin \varphi))$.

D.3.3 Camera to subject distance

D.3.3.1 Description

Camera to subject distance (CSD) should be obtained from the image metadata, e.g. from the JPEG image EXIF SubjectDistance (ID=37382, 0x9206) tag value as per the EXIF standard 2.3, JEITA CP-3451C. ISO/IEC 39794-5:2019, Table D.2 places the followint requirements on CSD:

— 0.7 m ≤ CSD ≤ 4 m for one-to-one applications, and

— 1 m ≤ CSD ≤ 4 m for one-to-many applications.

D.3.3.2 Computation of the native quality measure

1) If EXIF is available, recover d, the camera to subject distance (EXIF SubjectDistance tag ID=37382, 0x9206) in m.

2) Otherwise, if EXIF or crop factor is not available, assume photographers used a standard and set d = 1.5 m.

D.3.3.3 Mapping the computation result to the target range of the quality component

1) Compute a conformance, $p_S = \mathrm{SIGMOID}(d, 0.3, 0.1)$ that penalizes short ranges.

2) Compute a long range conformance, $p_L = \mathrm{SIGMOID}(-d, -10, 4)$ that penalizes long ranges.

3) Compute the product $Q = \mathrm{ROUND}(100\, p_S p_L)$.

D.3.4 Motion blur prevention

Motion blur is one mechanism by which resolution is reduced. It can often be quickly remediated by asking the subject to be still, or by guiding the photographer to use shorter integration times and more light.

Annex E
(informative)

OFIQ testing reports

Performance of the OFIQ implementation is documented in the OFIQ developer's report.[61]

Some performance aspects of OFIQ, prior editions of OFIQ, and private sector developers' QAAs, appear in a NIST FATE SIDD report.[62]

Some analysis of demographic variability of the OFIQ measures appears in References [62] and [63].

Annex F
(informative)

Guidance for sequential use of ISO/IEC 29794-5 quality components

If the unified quality score of a face image is low, it can be appropriate to inspect the defects of the image automatically based on the quality vector to give actionable feedback. It is advised to inspect the defects in the order as listed in Table F.1. The order has been derived from the logical dependencies between quality components as implemented in OFIQ. It is advised to abort further inspection as soon as the first quality component value suggests a defect of that component. For the first defect detected, corresponding actionable feedback can be given.

Table F.1 — Suggested order of inspecting quality components

#	Quality component	Interpretation for actionable feedback or adjustments in production
1	Single face present	No unique subject detected
2	Head pose angle yaw frontal alignment	Subject not posed frontal to the camera
3	Head pose angle pitch frontal alignment	Subject not posed frontal to the camera
4	Head pose angle roll frontal alignment	Subject not posed frontal to the camera
5	Inter-eye distance	Subject too far or too close to the camera
6	Head size	Subject too far or too close to the camera
7	Leftward crop of face in image	Subject’s face is potentially not well-centred on the acquired image
8	Rightward crop of face in image	
9	Margin above face in image	
10	Margin below face in image	
11	Face occlusion prevention	Subject’s face occluded
12	Eyes visible	Subject’s eye or eyes occluded
13	Mouth occlusion prevention	Subject’s mouth occluded
14	Under-exposure prevention	Dark lighting conditions or too short exposure duration
15	Over-exposure prevention	Bright lighting conditions or too long exposure duration
16	Illumination uniformity	Non-uniform lighting
17	Compression artefact prevention	Compression level too high
18	Sharpness	Out of focus or, camera or object moving or insufficient resolution
19	Mouth closed	Subject’s mouth open
20	Eyes open	Subject’s eyes closed
21	Natural colour	Non-neutral lighting colour
22	Moments of the luminance distribution	Bad lighting conditions
23	Dynamic range	Bad lighting conditions
24	Expression neutrality	Subject’s expression not neutral
25	Background uniformity	Background not uniform
26	No head covering	Subject is wearing a head cover

Bibliography

[1] Patrick Grother, Elham Tabassi, *Performance of Biometric Quality Measures*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 29, Issue 4, April 2007. https://ieeexplore.ieee.org/document/4107559/

[2] Phillips P.J., Beveridge J.R., Bolme D., Draper B.A., Givens G.H., Lui Y.M. et al. *On the Existence of Face Quality Measures*, In proc. IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS 2013), September 30, 2013. https://ws680.nist.gov/publication/get_pdf.cfm?pub_id=914258

[3] Han Y., Cai Y., Cao Y., Xu X. *Monotonic Regression: A New Way for Correlating Subjective and Objective Ratings in Image Quality Research,* IEEE Transactions on Image Processing, Volume: 21, Issue: 4, April 2012, Page(s): 2309 – 2313, 06 October 2011, DOI: 10.1109/TIP.2011.2170697

[4] Ji H., Liu C. *Motion blur identification from image gradients*. 2008 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2008.

[5] ISO 11664-4:2008, *(CIE S 014-4/E:2007) Colorimetry -- Part 4: CIE 1976 L*a*b* Colour space*

[6] Chong-Yaw Wee and Raveendran Paramesran. *Measure of image sharpness using eigenvalues.* Information Sciences 177.12 (2007): 2533-2552.

[7] Weber F. *Some quality measures for face images and their relationship to recognition performance*, NIST Biometric Quality Workshop I, 2006. [PDF]

[8] Beveridge J.R., Givens G.H., Phillips P.J., Draper B.A., Lui Y.M. *Face Region In Focus Measure (FRIFM) - Focus on quality, predicting FRVT 2006 performance*. In Proceedings of the Eighth International Conference on Automatic Face and Gesture Recognition, 2008.

[9] Grimmer M., Rathgeb C., Veldhuis R., Busch C. NeutrEx: A 3D Quality Component Measure on Facial Expression Neutrality, in proc. International Joint Conference on Biometrics (IJCB 2023).

[10] E. Tabassi M. Olsen. O. Bausinger, C. Busch, A. Figlarz, G. Fiumara, O. Henniger, J. Merkle, T. Ruhland, C. Schiel, M. Schwaiger *NISTI R 8382 NFIQ 2 NIST Fingerprint Image Quality*https://doi.org/10.6028/NIST.IR.8382

[11] ISO/IEC 19785-2:2021, *Information technology — Common Biometric Exchange Formats Framework — Part 2: Biometric registration authority*

[12] Openvino: https://github.com/openvinotoolkit/open_model_zoo/blob/master/models/intel/index.md

[13] D.E. King: Dlib-ml: A machine learning toolkit. Journal of Machine Learning Research, 10:1755–1758, 2009. Available from http://dlib.net

[14] Mediapipe: https://google.github.io/mediapipe/solutions/face_mes

[15] Deng J., Guo J., Ververas E., Kotsia I., Zafeiriou S. RetinaFace: Single-shot multi-level face localisation in the wild. In IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020

[16] Y. Zhu, H. Cai, S. Zhang, C. Wang, Y. Xiong: TinaFace: Strong but Simple Baseline for Face Detection. 2021

[17] X. Guo, S. Li, J. Yu, J. Zhang, J. Ma, L. Ma, W. Liu, H. Ling: PFLD: A Practical Facial Landmark Detector. 2021

[18] K. Zhang, Z. Zhang, Z. Li, Y. Qiao: Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks. IEEE Signal Processing Letters, 23(10):1499–1503, 2016

[19] Bazarevsky V., Kartynnik Y., Vakunov A., Raveendran K., Grundmann M. BlazeFace: Sub-millisecond Neural Face Detection on Mobile GPUs, July 2019, [1)]

1) https://arxiv.org/abs/1907.05047

[22] Xiao K., Yates J.M., Zardawi F., Sueeprasan S., Liao N., Gill L. et al. Characterising the variations in ethnic skin colours: a new calibrated data base for human skin. Skin Research and Technology 2017; 23: 21–29. Published by John Wiley & Sons Ltd, 2016 https://onlinelibrary.wiley.com/doi/pdf/10.1111/srt.12295

[23] Wang M., Xiao K. Sophie Wuerger, Vien Cheung, Ming Ronnier Luo, Measuring H.S.C. 23rd Color and Imaging Conference Final Program and Proceedings, Society for Imaging Science and Technology, 2015 https://eprints.whiterose.ac.uk/91398/8/wang_xiao_wuerger_cheung_luo_CIC2015_final.pdf

[24] Wang M. The influence of the skin colour on the perceived attributes, The University of Leeds, 2017 https://etheses.whiterose.ac.uk/19342/1/Mengmeng%20PhD%20thesis.pdf

[25] Ibsen M., Rathgeb C., Fink T., Drozdowski P., Busch C. Impact of facial tattoos and paintings on face recognition systems, IET Biometrics, 2021 https://ietresearch.onlinelibrary.wiley.com/doi/pdfdirect/10.1049/bme2.12032

[27] Bahmani K. et al. "SREDS: A dichromatic separation based measure of skin color." 2021 IEEE International Workshop on Biometrics and Forensics (IWBF). IEEE, 2021.

[28] W. Funk, M. Arnold, C. Busch, A. Munde: "Evaluation of Image Compression Algorithms for Fingerprint and Face Recognition Systems" in Proceedings from the sixth IEEE Systems, Man Cybernetics (SMC), (2005)

[29] Schlett T., Schachner S., Rathgeb C., Tapia J., Busch C. "Effect of Lossy Compression Algorithms on Face Image Quality and Recognition", in Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, GR, June 4-10, (2023)

[30] SSD face detector CNN available from https://github.com/sr6033/face-detection-with-OpenCV-and-DNN

[31] Wu W., Qian C., Yang S., Wang Q., Cai Y., Zhou Q. "Look at Boundary: A Boundary-Aware Face Alignment Algorithm". Proceedings of IEEE CVPR 2018 https://wywu.github.io/projects/LAB/WFLW.html

[32] face3d0725 CNN model. Available from https://github.com/face3d0725/FaceExtraction

[33] Yin X., Chen L. "FaceOcc: A diverse, high-quality face occlusion dataset for human face extraction." arXiv preprint arXiv:2201.08425 (2022).

[34] Yin X. Huang, Di, Fu, Zehua, Wang, Yunhong, and Chen, Liming. "Segmentation-Reconstruction-Guided Facial Image De-Occlusion". 17th IEEE Intl. Conference on Automatic Face and Gesture Recognition 2023 (FG'2023), Hawaiii, USA, 2023.

[35] Yin X. Huang, Di, and Chen, Liming, "Non-Deterministic Face Mask Removal Based on 3D Priors". 2022 IEEE International Conference on Image Processing (ICIP), Bordeaux, France, 2022.

[36] Martinez A.M., Benavente R. "The AR Face Database. CVC Technical Report #24," 1998.

[37] Gross R., Matthews I., Cohn J., Kanade T., Baker S. "Multi-PIE," in *8th IEEE International Conference on Automatic Face & Gesture Recognition*, Amsterdam, Netherlands, 2008.

[38] J. Phillips, H. Wechsler, J. Huang and P. J. Rauss, "The FERET database and evaluation procedure for face-recognition algorithms," *Image and Vision Computing,* vol. 16, no. 5, pp. 295-306, 1998.

[39] Burgos-Artizzu X.-P., Perona P., Dollar P. "Robust face landmark estimation under occlusion," in *International Conference on Computer Vision (ICCV)*, 2013.

[40] X. Yin and L. Chen, "FaceOcc: A Diverse, High-quality Face Occlusion," in *arXiv preprint*, 2022.

[41] S. M. Mavadati, M. H. Mahoor, K. Bartlett, P. Trinh and J. F. Cohn, "DISFA: A spontaneous facial action intensity database," *IEEE Transactions on Affective Computing,* vol. 4, no. 2, pp. 151-160, 2013.

[42] W. Gao, B. Cao, S. Shan, D. Zhou, X. Zhang and D. Zhao, "The CAS-PEAL Large-Scale Chinese Face Database and Baseline Evaluations," 2004.

[43] Lee C.-H., Liu Z., Wu L., Luo P. "MaskGAN: Towards Diverse and Interactive Facial Image Manipulation," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 2020.

[44] U. Cheema and S. Moon, "Sejong face database: A multi-modal disguise face database," *Computer Vision and Image Understanding*, 2021.

[45] IEEE 754, *Standard for Floating-Point Arithmetic, 2019,*[2)]

[46] ISO/IEC 14496-2:2004, *Information technology — Coding of audio-visual objects — Part 2: Visual*

[47] Report I.T.U.-R.B.T. 2446-0 (04/2019), Methods for conversion of high dynamic range content to standard dynamic range content and vice-versa https://www.itu.int/dms_pub/itu-r/opb/rep/R-REP-BT.2446-2019-PDF-E.pdf

[48] Recommendation ITU-R BT.2100-2 (07/2018), Image parameter values for high dynamic range television for use in production and international programme exchange PG Uused

[49] Guo J., Zhu X., Yang Y., Yang F., Lei Z., Stan Z. Li, Towards Fast, Accurate and Stable 3D Dense Face Alignment, ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIX Aug 2020 Pages 152–168 https://doi.org/10.1007/978-3-030-58529-7_10

[50] The official PyTorch implementation of Towards Fast, Accurate and Stable 3D Dense Face Alignment, ECCV 2020. https://github.com/cleardusk/3DDFA_V2

[51] [HSE] https://github.com/HSE-asavchenko/face-emotion-recognition

[52] [EENE] https://github.com/dasec/Efficient-Expression-Neutrality-Estimation

[53] [ADNET] https://github.com/huangyangyu/ADNet

[53] [PARSE] https://github.com/zllrunning/face-parsing.PyTorch

[54] https://github.com/IrvingMeng/MagFace trained on MS1MV2

[55] Meng Q., Zhao S., Huang Z., Zhou F. MagFace: A universal representation for face recognition and quality assessment. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR),2021. Available fromhttps://doi.org/10.1109/CVPR46437.2021.01400

[56] Deng J., Guo J., Xue N., Zafeiriou S. ArcFace: Additive angular margin loss for deep face recognition. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR),2019. Available fromhttps://doi.org/10.1109/CVPR.2019.00482

[57] ISO/IEC 29794-5, *Face Image Quality – Maintenance Portal,*[3)]

[58] ISO/IEC 29794-5, *Face Image Quality – Supporting Models,*[4)]

[59] Bundesamt für Sicherheit in der Informationstechnik (BSI), "Open Face Image Quality – Source Code"https://github.com/BSI-OFIQ/OFIQ-Project.

[60] Bundesamt für Sicherheit in der Informationstechnik (BSI), "Open Face Image Quality – Homepage"https://bsi.bund.de/dok/OFIQ-e

[61] Bundesamt für Sicherheit in der Informationstechnik (BSI), "Open Face Image Quality – Performance Report"[5)]

[62] NIST Face Analysis Technology Evaluation – Face Image Quality Specific Image Defect Detectionhttps://pages.nist.gov/frvt/html/frvt_quality.html

2) https://standards.ieee.org/ieee/754/6210/

3) https://standards.iso.org/iso-iec/29794/-5/ed-1/en/

4) https://standards.iso.org/iso-iec/29794/-5/ed-1/en/OFIQ-MODELS.zip

5) https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/OFIQ/Projektabschlussbericht_OFIQ_1_0.pdf

[63] Idemia, “OFIQ Evaluation Report”, August2024, https://www.idemia.com/wp-content/uploads/2024/08/ofiq-evaluation-demographic-factors-072024.pdf

[64] EU Commission Implementing Decision2019/329, 25 February 2019, “Specifications for the quality, resolution and use of fingerprints and facial image for biometric verification and identification in the Entry/Exit System(EES)”

[65] International Civil Aviation Organization (ICAO) – Document 9303 https://www.icao.int/publications/documents/9303_p3_cons_en.pdf

[68] [face-parsing.PyTorch] https://github.com/VisionSystemsInc/face-parsing.PyTorch

[69] [CelebAMask-HQ] https://github.com/switchablenorms/CelebAMask-HQ

[70] Face Alignment Across Large Poses: A 3D Solution, http://www.cbsr.ia.ac.cn/users/xiangyuzhu/projects/3DDFA/main.htm

ICS 35.240.15

Price based on 63 pages

iso.org