



DPC: 24/30423099 DC

BSI Group Headquarters

389 Chiswick High Road London W4 4AL

Tel: +44 (0)20 8996 9000

Fax: +44 (0)20 8996 7400

www.bsigroup.com

Date: 03 April 2024

Origin: International

Latest date for receipt of comments: 27 May 2024

Project No. 2020/01745

Responsible committee: IST/44/-/3 Biometric data interchange formats

Interested committees:

Title: Draft BS ISO/IEC 29794-5 Information technology - Biometric sample quality. Part 5: Face image data.

Please notify admin.start@bsigroup.com if you are aware of any keywords that might assist in clarifying or identifying the standard or of the content of this standard:

- i) has any issues related to 3rd party IPR, patent or copyright
- ii) affects other national standard(s)
- iii) requires additional national guidance or information

WARNING: THIS IS A DRAFT AND MUST NOT BE REGARDED OR USED AS A BRITISH STANDARD. THIS DRAFT IS NOT CURRENT BEYOND 27 May 2024

This draft is issued to allow comments from interested parties; all comments will be given consideration prior to publication. No acknowledgement will normally be sent. **See overleaf for information on the submission of comments.**

No copying is allowed, in any form, without prior written permission from BSI except as permitted under the Copyright, Designs and Patent Act 1988 or for circulation within a nominating organization for briefing purposes. Electronic circulation is limited to dissemination by e-mail within such an organization by committee members.

Further copies of this draft may be purchased from BSI Shop <http://shop.bsigroup.com> or from BSI Customer Services, Tel: +44(0) 20 8996 9001 or email cservices@bsigroup.com. British, International and foreign standards are also available from BSI Customer Services.

Information on the co-operating organizations represented on the committees referenced above may be obtained from <https://standardsdevelopment.bsigroup.com/>.

E-mail: admin.start@bsigroup.com

Introduction

This draft standard is based on national and international discussions. Your comments on this draft are invited and will assist in the preparation of the consequent standard.

For international standards, comments will be reviewed by the relevant UK national committee before sending the consensus UK vote and comments to the international committee, which will then decide appropriate action. If the international standard is approved, it is usual for the text to be published as a British Standard.

For national standards, comments will be reviewed by the relevant UK national committee and the resulting standards published as a British Standard.

UK Vote

Please indicate whether you consider the UK should submit a negative (with supporting technical reasons) or positive vote on this draft. Please indicate if you are aware of any reason why this draft standard should not be published as a British Standard.

Submission of Comments

- **Annotated drafts are not acceptable and will be rejected.**
- All comments should be submitted online at <https://standardsdevelopment.bsigroup.com/>. You will need to register in order to comment.

Template for comments and secretariat observations

Date: xx/xx/20xx	Document: ISO/DIS xxxx
------------------	------------------------

1	2	(3)	4	5	(6)	7
MB	Clause No./ Subclause No./Annex (e.g. 3.1)	Paragraph/ Figure/Table/Note	Type of comment	Comment (justification for change) by the MB	Proposed change by the MB	Secretariat observations on each comment submitted
	3.1	Definition 1	ed	Definition is ambiguous and needs clarifying.	Amend to read '...so that the mains connector to which no connection...'	
	6.4	Paragraph 2	te	The use of the UV photometer as an alternative cannot be supported as serious problems have been encountered in its use in the UK.	Delete reference to UV photometer.	



DRAFT International Standard

ISO/IEC DIS 29794-5

Information technology — Biometric sample quality —

Part 5: Face image data

ICS: 35.240.15

ISO/IEC JTC 1/SC 37

Secretariat: **ANSI**

Voting begins on:
2024-04-03

Voting terminates on:
2024-06-26

This document is circulated as received from the committee secretariat.

THIS DOCUMENT IS A DRAFT CIRCULATED FOR COMMENTS AND APPROVAL. IT IS THEREFORE SUBJECT TO CHANGE AND MAY NOT BE REFERRED TO AS AN INTERNATIONAL STANDARD UNTIL PUBLISHED AS SUCH.

IN ADDITION TO THEIR EVALUATION AS BEING ACCEPTABLE FOR INDUSTRIAL, TECHNOLOGICAL, COMMERCIAL AND USER PURPOSES, DRAFT INTERNATIONAL STANDARDS MAY ON OCCASION HAVE TO BE CONSIDERED IN THE LIGHT OF THEIR POTENTIAL TO BECOME STANDARDS TO WHICH REFERENCE MAY BE MADE IN NATIONAL REGULATIONS.

RECIPIENTS OF THIS DRAFT ARE INVITED TO SUBMIT, WITH THEIR COMMENTS, NOTIFICATION OF ANY RELEVANT PATENT RIGHTS OF WHICH THEY ARE AWARE AND TO PROVIDE SUPPORTING DOCUMENTATION.

Reference number
ISO/IEC DIS 29794-5:2024(en)

Full document details: 24/30423099 DC

 **Documents**
Document Sharing Hub
© ISO/IEC 2024



COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2024

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

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

Published in Switzerland

© ISO/IEC 2024 – All rights reserved

Contents

Page

Foreword	v
Introduction	vi
1 Scope	1
2 Normative references	1
3 Terms and definitions	2
4 Symbols and abbreviated terms	4
5 Conformance	4
6 Common computations	6
6.1 Overview	6
6.2 Conversion of 16 bits per channel images to 8 bits per channel images	6
6.3 Conversion of high bit-depth images to 8 bit greyscale or 24 bit colour images	6
6.4 Face detection	6
6.5 Face landmark estimation	8
6.6 Landmarked region segmentation	10
6.7 Face alignment	10
6.8 Face parsing	11
6.9 Face occlusion segmentation	12
6.10 Computing eye centres and inter-eye distance	13
6.11 Pose estimation	14
6.12 Conversion of 8-bits-per-channel colour images to luminance	16
6.13 Conversion of 8-bits-per-channel colour images to CIELAB space	16
6.14 Handling of greyscale images	17
6.15 Luminance histogram	17
6.16 Entropy	18
6.17 Expressing binary quantities as continuous values	18
6.18 Representation and arithmetic of real and integer numbers	18
6.19 Normalization of image colour values	19
7 Quality measures	19
7.1 General	19
7.2 Quality score (unified)	19
7.2.1 Description	19
7.2.2 Computation	20
7.2.3 Mapping the computation result to the target range of the quality component	20
7.3 Capture-related quality components	21
7.3.1 General	21
7.3.2 Background uniformity	21
7.3.3 Illumination uniformity	22
7.3.4 Moments of the luminance distribution	23
7.3.5 Under-exposure prevention	24
7.3.6 Over-exposure prevention	25
7.3.7 Dynamic range	26
7.3.8 Sharpness	27
7.3.9 No compression artifacts	28
7.3.10 Natural colour	29
7.4 Subject-related quality components	30
7.4.1 General	30
7.4.2 Single face present	30
7.4.3 Eyes open	31
7.4.4 Mouth closed	32
7.4.5 Eyes visible	33
7.4.6 Mouth occlusion prevention	33
7.4.7 Face occlusion prevention	34

7.4.8	Inter-eye distance.....	35
7.4.9	Head size.....	36
7.4.10	Crop of the face image	36
7.4.11	Pose	38
7.4.12	Expression neutrality	40
7.4.13	No head covering.....	41
8	Face image quality block	42
8.1	Binary encoding	42
8.2	XML encoding	42
8.3	Organization identifiers.....	42
8.4	Algorithm identifiers	42
Annex A (normative) Conformance test assertions		44
Annex B (informative) Quantitative goal for face image QAAs		49
Annex C (informative) Applications of quality values		52
Annex D (Informative) Quality requirements with no quality measure		54
Annex E (Informative) OFIQ Testing Results.....		57
Annex F (informative) Guidance for sequential use of 29794-5 quality components		66
Bibliography		67

Foreword

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

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

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

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

For an explanation 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.

This standard does not include contractual, legal or statutory requirements. Voluntary standards do not replace national laws, with which standards users are understood to comply and which take precedence. Local regulations / regional laws can apply.

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

The quality measures included in this part of 29794 were selected because 39794-5 includes guidance on how they should be controlled. The implementations of those quality measures were evaluated for performance and have license conditions that allow their integration in commercial systems (i.e. products).

hyperlinks:

1. “allow” to <https://bsi.bund.de/dok/OFIQ-e>
2. “evaluated” to https://pages.nist.gov/frvt/html/frvt_quality.html

A list of all parts in the ISO/IEC 29794 series can be found on the ISO website. This second edition cancels and replaces the first edition (ISO/IEC TR 29794-5:2010), which has been completely revised to become an International Standard.

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.

Introduction

Even though face recognition error rates dropped massively with the introduction of algorithms based on deep learning, they remain significant, depending on the imaging process, human factors design, level of biometric capture subject cooperation, the particular comparison algorithm, and associated threshold and decision logic. This standard is needed because without significant modernization of capture procedures, recognition errors will become more prevalent as volumes increase. Drivers for improved capture are as follows.

- **[Need for improved useability]** As biometric systems, in general, have improved, it has become clear that improved useability 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 may see 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 for immigration exit 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. This contrasts with the situation with fingerprint and iris biometrics where capture devices run in an auto-capture quality-assessment loop, with explicit awareness of the kind of image that should be collected.
- **[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 may 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 eyewear. 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, 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.

This standard establishes requirements on software that inspects exactly one image. This standard does not establish requirements on software that compares two or more images (such as biometric recognition).

[Table 1](#) expresses that quality issues due to mis-presentation (often associated with human factors design) and those 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 — ... Other static characteristics — Thick or dark glasses — Permanent jewellery — Makeup and cosmetics — ...	Capture process and capture device properties: — image resolution — optical distortions — sub-optimal camera angle — ... 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 — ...	Scenery — background moving objects — variation in lightning Capture device variation — de-focus — camera vibration — sub-optimal camera angle — poor exposure — ...

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 Annex D.1 for the use-case of collection of reference samples for ID documents, and two other use-cases defined in [clause 5](#). It 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 quality of pairs or sequences of images
- assessing quality of 3D captures,
- encodings of face image quality data, and
- performance evaluation of face image quality assessment algorithms.

NOTE 1 The use cases within scope of this document primarily address assessment of images from data capture subjects who consent to processing of their biometric data, or for whom biometric capture is operationally authorized. Use of this standard may be subject to local regulations.

NOTE 2 The computations of this document can be applied separately to each image in a pair or sequence.

NOTE 3 By defining image quality measurements, this document is intended to improve accuracy of automated face recognition systems. Improved quality can also improve human review of images.

NOTE 4 Some of the computations of this standard can be effective on images captured with illumination at non-visible wavelengths.

NOTE 5 Encoding of quality data is defined in ISO/IEC 29794-1.

NOTE 6 The methodology for performance evaluation of quality assessment algorithms is defined in ISO/IEC 29794-1.

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

IEC 61966-2-1/AMD 1:2003, *Multimedia systems and equipment — Colour measurement and management — Part 2-1: Colour management — Default RGB colour space — sRGB — Amendment 1*

RECOMMENDATION ITU-R BT, 2100-2 (07/2018), Image parameter values for high dynamic range television for use in production and international programme exchange

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

ISO/IEC 14496-2:2004, *Information technology — Coding of audio-visual objects — Part 2: Visual*

ISO/IEC 2382-37:2022, *Information technology — Vocabulary — Part 37: Biometrics*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 2382-37, ISO/IEC 39794-5:2019, 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 might be that given in ISO/IEC 39794-5:2019 Annex D.1 which specifies a reference face image for a machine-readable travel document.

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

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

[SOURCE: ISO/IEC 39794-5:2019, 3.27]

Note 1 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).

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 [clause 6.2](#).

Note 3 to entry: For face images that are compliant to ISO/IEC 39794-5 Annex D.1, the face bounding box corresponds to the “face portrait” as defined in ISO/IEC 39794-5 in clause 3.28.

3.6

face landmarks

A 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 [6.3](#)

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 landmark detection algorithm, as described in [clause 6.5](#) and encompassing the area from crown to chin and left ear to right ear.

Note 2 to entry: The inner region as defined in ISO/IEC 39794-5, term 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 comprise 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 clause 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

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 meant 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 e.g. Adobe RGB (1998) or ProPhoto RGB (ROMM RGB) are used as specified in ISO/IEC 39794-5, clause D.1.4.2.9.

3.14**quality assessment algorithm****quality algorithm**

algorithm to calculate a quality measure

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

Note 1 to entry: The ISO/IEC 19785 series of standards (CBEFF) uses the term "quality algorithm".

4 Symbols and abbreviated terms**Table 2 — Acronyms and abbreviations**

Acronym	Definition
BGR	color space with the order of colors (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
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	color space with the order of colors (red, green, blue)
RQM	raw quality measure
UC	use-case

5 Conformance

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

A face image quality assessment implementation is conformant to this part of ISO/IEC 29794 if it implements the computations marked as being mandatory (M) in [Table 3](#) for the claimed use-case(s).

EXAMPLE Developer X asserts that its algorithm Y is conformant to ISO/IEC 29794-5:2024 use-case 3.

A face image quality assessment algorithm conforms to this document if it conforms to the normative requirements of [clauses 6](#) and [7](#) and the requirements specified in [Annex A](#).

A face image quality block shall conform to this document if its structure and data values conform to the formatting requirements of [clause 8](#) (face image quality block) and its quality measures are computed using the methods specified in [clause 7](#).

ISO/IEC DIS 29794-5:2024(en)

Conformance to normative requirements of [clause 8](#) fulfils Level 1 and Level 2 conformance as specified in ISO/IEC 39794-1:2019, Annex C. Conformance to normative 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 might understand M as a recommendation to execute the software.

NOTE 2 The quality score is an image quality measure that describes the fitness of a biometric sample to accomplish the comparison decision and aligns to the requirement of ISO/IEC 29794-1. A quality component is a quality measure that can contribute to the computation of a (unified) quality score. For quality components in rows 2 to 36 of [Table 3](#), the “higher is better” need not apply.

NOTE 3 For standardized interchange of quality values produced by implementations of this document, ISO/IEC 29794-1 defines a standardised quality block and requires that quality measures shall 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 face image with instantaneous response.

Table 3 — Conformance requirements by activity

#	Face image quality measure	Sub-clause	UC1 (ID Docs.)	UC2 (System enrol- ment)	UC3 (Probe for immediate 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 artifacts	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	7.4.4	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	M	M	M
21.	Rightward crop of face in image	7.4.10.2	M	M	M
22.	Downward crop of face in image	7.4.10.3	M	M	M
23.	Upward crop of face in image	7.4.10.4	M	M	M

Table 3 (continued)

#	Face image quality measure	Sub-clause	UC1 (ID Docs.)	UC2 (System enrol- ment)	UC3 (Probe for immediate recognition)
24.	Pose angle yaw frontal alignment	7.4.11.2	M	M	0
25.	Pose angle pitch frontal alignment	7.4.11.3	M	M	0
26.	Pose angle roll frontal alignment	7.4.11.4	M	M	0
27.	Expression neutrality	7.4.12	M	0	0
28.	No head covering	7.4.13	M	0	0
29.	Radial distortion	D.2.1	0	0	0
30.	Pixel aspect ratio	D.2.2	0	0	0
31.	Gaze	D.3.1	0	0	0
32.	Shoulder presentation	D.3.2	0	0	0
33.	Camera subject distance	D.3.3	0	0	0
34.	Motion blur prevention	D.3.4	0	0	0

6 Common computations

6.1 Overview

[Clause 6](#) supports the computations in [clause 7](#) by defining certain steps that appear in more than one of its subclauses.

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

Conversion of 16-bit scRGB is defined in the IEC 61966-2.1 standard using conversions defined in the IEC 61966-2-2 annex A.2 Transformation from 16-bit scRGB values.

Clauses 3.1 and 4.1 of IEC 61966-2-2:2003, 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

The ISO/IEC 39794-5:2019 standard allows RAW format capture and more than 8 bits encoding as defined in D.1.4.2.4. RAW format can be a proprietary format with non-normative encoding specification. 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 done 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 the Report ITU-R BT.2446-0 (04/2019).

6.4 Face detection

This subclause supports the computations of [clause 7](#) that depend on detection of faces in a face image. The result of the face detection is the number of found faces and, for each face, its face bounding box.

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 a face image I represented in the BGR colour space with 8 bits per channel and uses the SSD face detector CNN^[30] as follows.

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

2. Pad I from the left and right side by $\lfloor 0.2w \rfloor$ and from the top and bottom by $\lfloor 0.2h \rfloor$, and update w and h to the new width and height of I .
3. Resize the image I to size (height, width) 300×300 using OpenCV with bilinear interpolation
4. Normalize the values of I using mean $\mu = (104.0, 117.0, 123.0)$ and standard deviation $\sigma = (1, 1, 1)$ 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 box table F of size $(N, 7)$, where each row of F contains the data of the detected face boxes:
 - a. Column 3 specifies the confidence
 - b. Columns 4-5 specify the relative (i.e. scaled to the interval $[0,1]$) x- and y-coordinates, respectively, of the upper left corner.
 - c. Columns 6-7 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. $F_{i,3} < 0.4$
 - b. The width of the face box is smaller than $1/20$ of the image width, i.e. $F_{i,6} - F_{i,4} < 0.05$
 - c. The face 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. Determine the row j with the largest face box area, i.e.,
$$j = \operatorname{argmax}((F_{i,7} - F_{i,5}) \cdot (F_{i,6} - F_{i,4})).$$

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

11. Obtain the coordinates (a, b) and (c, d) of the upper left corner and the lower right corner, respectively, as
$$a = \lfloor w \cdot F_{j,4} + 0.5 \rfloor, b = \lfloor h \cdot F_{j,5} + 0.5 \rfloor,$$

$$c = \lfloor w \cdot F_{j,6} + 0.5 \rfloor, d = \lfloor h \cdot F_{j,7} + 0.5 \rfloor.$$
12. Return the face box coordinates (a, b, c, d) .

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 the significant roll angle.

NOTE For cooperative, non-cooperative and unconstrained face recognition applications, the face bounding box and detection definitions here accommodate images where one or two eyes are visible, or may be closed or occluded. Furthermore, it accommodates images where multiple faces may be present.

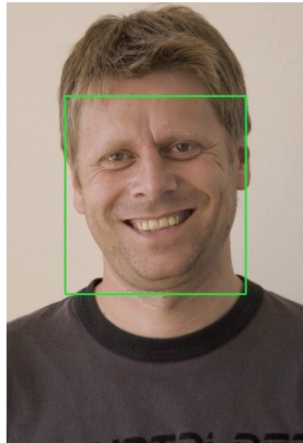


Figure 1 — Example face bounding box

6.5 Face landmark estimation

This subclause supports the computations of [clause 7](#), where many quality component algorithms either require specific landmarks or the landmarked region.

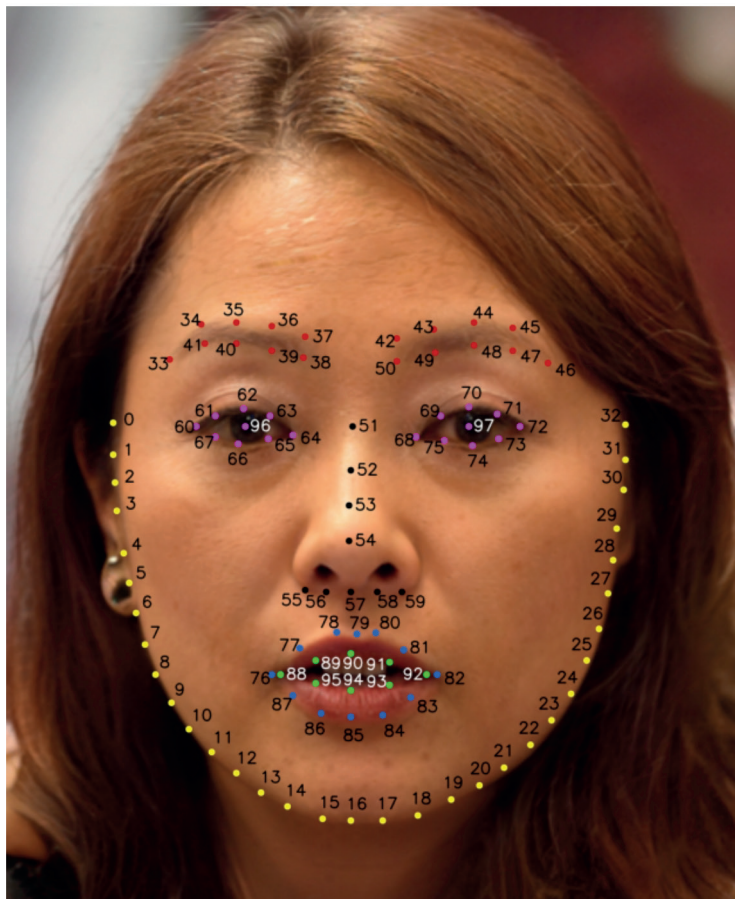


Figure 2 — Location and naming of landmarks in an example face image

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 CNN model published at <https://github.com/huangyangyu/ADNet> trained on the Wider Faces in the Wild dataset^[32] and takes as input an RGB image I (8 bits per channel) along with the face box (a,b,c,d) computed by the face detection algorithm specified in [clause 6.2](#). The output is a set of 98 landmarks – see the example in [Figure 2](#). These are used instead of those proposed in ISO/IEC 14496-2:2004, which were referred to in ISO/IEC 39794-5:2019, because there is no robust implementation for computing them.

Table D — Meaning of the [Figure 1](#) landmark colors

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

The computation of the facial landmarks takes as input the image and a face bounding box output by the face detection algorithm specified in [clause 6.2](#).

- Determine the height h and the width w of I .
- Extend the face box on both sides to square shape:
 - If $c-a < d-b$ then update $(a,c) \leftarrow \left(\left\lfloor \frac{a+c}{2} - \frac{d-b}{2} \right\rfloor, \left\lfloor \frac{a+c}{2} - \frac{d-b}{2} \right\rfloor \right)$
 - Otherwise, and if $c-a > d-b$ then update $(b,d) \leftarrow \left(\left\lfloor \frac{b+d}{2} - \frac{c-a}{2} \right\rfloor, \left\lfloor \frac{b+d}{2} - \frac{c-a}{2} \right\rfloor \right)$
- If necessary, pad the image and update the face box coordinates accordingly by setting
 - $p_{\text{left}} = \max(0, -a)$,
 - $p_{\text{right}} = \max(0, c - w + 1)$,
 - $p_{\text{top}} = \max(0, -b)$,
 - $p_{\text{bottom}} = \max(0, d - h + 1)$.
- Pad the image from left, top, right, bottom side by p_{left} , p_{top} , p_{right} , p_{bottom} pixels, respectively.
- Update a, b, c, d to $a + p_{\text{left}}$, $b + p_{\text{top}}$, $c + p_{\text{right}}$, $d + p_{\text{bottom}}$
- Crop I to (a, b, c, d) , i.e. to $I[a:c, b:d]$
- Resize I to size (256,256)
- Normalize the values of I using mean $\mu = 127.5$ and standard deviation $\sigma = 127.5$ to obtain the array A .
- Encode A as input tensor $T^{[0]}$ of dimensions (1,256,256).
- Run a forward pass through the model using $T^{[1]}$ as input to obtain the output tensor $T^{[1]}$ of dimensions (1,98,2).
- Reshape $T^{[1]}$ to an array A of dimensions (98,2).
- De-normalize A to obtain the array B specifying the landmark positions within the cropped and resized image by setting

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

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

$$x_i = B_{i,1} \frac{d-b}{256} + a - p_{\text{left}} \text{ and } y_i = B_{i,2} \frac{d-b}{256} + b - p_{\text{top}}.$$

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

6.6 Landmarked region segmentation

The algorithm takes as input a set of landmarks, either the original landmarks computed by the algorithm in [clause 6.5](#) or the transformed landmarks computed by the algorithm in [clause 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 .

NOTE 1 An example of the segmented map appears in [Figure 3](#).

NOTE 2 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 and corresponding facial landmarks and then performs the following steps:

1. For both eyes, compute the eye centres C_r , C_l as the centre of mass of the landmarks marking the canthi of the eyes.
2. Determine the position N of the nose tip.
3. Determine the positions M_r and M_l of the subject's right and left mouth corner.
4. Determine the isogonal affine transform that optimizes the mapping of the points C_r , C_l , N , M_r , and M_l to the target points (251, 272), (364, 272), (308, 336), (262, 402), and (355, 402), respectively. Therefore, the OpenCV (version 4.5.5) function `estimateAffinePartial2D` should be used with the 'least median square' approach (method=LMEDS).
5. Apply the transformation T to the image and crop and/or pad the image to dimension 616x616. For padding use black color (0,0,0).
6. Apply the transformation T to the landmarks.
7. Output the transformed image, the transformed landmarks and the transformation matrix T .

6.8 Face parsing

Face parsing supports some computations in [clause 7](#) by generating a segmentation map distinguishing the following 19 classes.



Figure 5 — Example segmentation map.

Table 5 — Label values with their feature or region

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 [Table 5](#) features and regions 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 RGB image (8 bits per channel) that has been aligned as specified in [clause 6.7](#) and uses the face3d0725 CNN model^[33-35].

1. Crop the image by 30 pixels from both sides and by 60 pixels from the bottom
2. Scale the image to 400x400 pixels 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).
4. Reshape the image array 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 400x400.
7. Return *S*.

NOTE 1 For UC1 it is a normative requirement in ISO/IEC 39794-5 Annex D.1.5.4 that only down-sampling is 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 done for quality assessment.

NOTE 2 Since the face parsing CNN has been trained on aligned images, the images should be aligned before using the CNN.

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 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](#). The algorithm takes as input an RGB image (8 bits per channel) that has been aligned as specified in [clause 6.7](#) and uses the face3d0725 CNN model^[32]:

1. Crop the image from all sides by 96 pixels.
2. Scale the image to 224x224 pixels 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 parsing CNN to obtain an output tensor of size (1,224,224).
6. Remove the first dimension of the output tensor to obtain a segmentation mask M of size (224,224).
7. Set all positive values of the segmentation mask M to 1, and all other values to 0.
8. Resize the segmentation mask M to 424x424 pixels using nearest neighbour interpolation (using truncation for rounding non-integer indices).
9. Pad the segmentation mask M by 96 pixels from all sides setting the padded regions to 0 indicating absence of face or occluded face.
10. Output the segmentation mask 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 $L_0, \dots, L_{97}$ estimated by the algorithm in [clause 6.5](#) and the head pose angle for the yaw, ϕ_{yaw} , computed by the algorithm in [clause 6.11](#).

1. The coordinates of the centre of the subject's left eye, (X_L, Y_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. $(X_L, Y_L) = (L_{68} + L_{72})/2$.

2. The coordinates of the centre of the subject's right eye, (X_R, Y_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. $(X_R, Y_R) = (L_{60} + L_{64})/2$.
3. If $|\cos \phi_{\text{yaw}}| < \varepsilon$ then output (X_L, Y_L) , (X_R, Y_R) and NaN and terminate the algorithm. Here ε denotes the tolerance value $1e-6$ and NaN is an abstract value that could not be computed (not a number).
4. Compute the inter-eye distance, IED , as the yaw-corrected Euclidean distance between the left and right eye centres:

$$IED = \frac{1}{\cos(\phi_{\text{yaw}})} \sqrt{(X_L - X_R)^2 + (Y_L - Y_R)^2}$$

5. Output (X_L, Y_L) , (X_R, Y_R) and IED

6.11 Pose estimation

An implementation shall be capable of estimating three head orientation angles (pitch, yaw 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](#).

The following algorithm uses a CNN model M published at https://github.com/cleardusk/3DDFA_V2 to estimate the head pose of a subject's face. The implementation estimates three head orientation angles (yaw, pitch and roll) from a face image.

The algorithm takes as input an image I in BGR colour format (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 box (a, b, c, d) is not within the image region $[0, w-1] \times [0, h-1]$, where w, h are the width and height of I , respectively, pad the image:
 - 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 120×120 using the OpenCV function *resize* with bilinear interpolation.
7. Normalize I using mean $\mu = 127.5$ and standard deviation $\sigma = 128$ to obtain the array A , i.e. set

$$A_{i,j,k} = \frac{I_{i,j,k} - \mu}{\sigma}$$

for all i, j, k .

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 $(1, 62)$.