ISO/IEC JTC 1/SC 37 N 3610

2010-02-11

ISO/IEC JTC 1/SC 37
Biometrics

Document Type: FCD text

Document Title: Text of 2nd FCD 19794-5, Biometric data interchange formats – Part 5: Face image data

Source: Project Editor

Document Status: As per Singapore recommendation 3.4, this document is circulated to SC 37 National Bodies for a four month letter ballot.

Action ID: Vote

Due Date: **2010-06-03**

No. of Pages: 125

Secretariat, ISO/IEC JTC 1/SC 37, American National Standards Institute, 25 West 43rd Street, New York, NY 10036; Telephone: 1 212 642 4932; Facsimile: 1 212 840 2298; Email: lrajchel@ansi.org

ISO/IEC 2nd FCD 19794-5	
Date: **2010-02-11**	Reference number: ISO/IEC JTC 1/SC 37 **N 3610**
Supersedes document	

THIS DOCUMENT IS STILL UNDER STUDY AND SUBJECT TO CHANGE. IT SHOULD NOT BE USED FOR REFERENCE PURPOSES.

ISO/IEC JTC 1/SC 37 Biometrics Secretariat: USA (ANSI)	Circulated to P- and O-members, and to technical committees and organizations in liaison for: - discussion at - comment by **- voting by (P-members only)** **2010-06-03** Please return all votes and comments via the online balloting system.

ISO/IEC JTC 1/SC 37

Title: Text of 2nd FCD 19794-5, Biometric data interchange formats – Part 5: Face image data

Project: 1.37.19794.5

Introductory note:

As per Singapore resolution 3.4, this document is circulated to SC 37 National Bodies for a four month letter ballot.

Recipients of this document are invited to submit notification of any relevant patent rights of which they are aware and to provide supporting documentation.

ISO/IEC JTC 1/SC 37 N 3610

Date: 2009-02-15

ISO/IEC WD 19794-5

ISO/IEC JTC 1/SC 37/WG 3

Secretariat: ANSI

Information technology – Biometrics – Biometric data interchange formats – Part 5: Face image data

Document type: International Standard
Document subtype:
Document stage: (20) Preparatory
Document language: E

Contents

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.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights.

ISO/IEC 19794-5 was prepared by Technical Committee ISO/TC JTC 1, Subcommittee SC 37, Biometrics.

This second edition cancels and replaces the first edition.

ISO/IEC 19794 consists of the following parts, under the general title *Biometric Data Interchange Formats — Part 5: Face Image Format for Data Interchange*:

— Biometric Data Interchange Formats:

— Part 1: Framework

— Part 2: Finger Minutiae Data

— Part 3: Finger Pattern Spectral Data

— Part 4: Finger Image data

— Part 5: Face Image data

— Part 6: Iris Image data

— Part 7: Signature/Sign Behavioural Data

— Part 8: Finger Pattern Skeletal Data

— Part 9: Vascular Image data

— Part 10: Hand Geometry Silhoutte Data

— Part 11: Signature/Sign Processed Dynamic Data

— Part 13: Voice Data

— Part 14: DNA Data

Introduction

Face images, also commonly referred to as displayed portraits, have been used for many decades to verify identity of persons. In recent years, digital face images are used in many applications including human examination as well as computer automated face recognition. Although photographic formats have been standardized in some cases such as passport and driver license, it is also demanded to define a standard data format of digital face images to allow interoperability among vendors.

This part of ISO/IEC 19794 is intended to provide a face image format for face recognition applications requiring exchange of face image data. The typical applications are:

1) human examination of facial images with sufficient resolution to allow a human examiner to ascertain small features such as moles and scars that might be used to verify identity;

2) human verification of identity by comparison of persons against facial images;

3) computer automated face biometric identification (one-to-many searching);

4) computer automated face biometric verification (one-to-one comparison).

To enable many applications on a variety of devices, including devices that have the limited resources required for data storage, and to improve face recognition accuracy, this part of ISO/IEC 19794 specifies not only a data format, but also scene constraints (lighting, pose, expression etc), photographic properties (positioning, camera focus etc), digital image attributes (image resolution, image size etc).

Several image types are introduced to define categories that satisfy requirements of some applications. Each requirement is specified for each image type.

This is a revision of ISO/IEC 19794-5:2004. The structure of the data format is not compatible with the previous version.

NOTE this part of ISO/IEC 19794 relies on other ISO standards.

Information technology – Biometrics – Biometric data interchange formats – Part 5: Face image data

1 Scope

1.1 General

This part of ISO/IEC 19794

— specifies a record format for storing, recording, and transmitting the information from one or more facial images or a short video stream of facial images, specifies scene constraints of the facial images,

— specifies photographic properties of the facial images,

— specifies digital image attributes of the facial images,

— provides best practices for the photography of faces.

1.2 Face Image Types

Each requirement is specified for the following Face Image Types, respectively.

— **Basic:** This is the fundamental Face Image Type that specifies a record format including header and representation data. All Face Image Types adhere to the properties of this type. No mandatory scene, photographic and digital requirements are specified for this image type.

— **Frontal:** A Basic Face Image Type that adheres to additional requirements appropriate for frontal face recognition and/or human examination. Two types of Frontal Face Image Types are defined in this document, Full Frontal and Token Frontal (or simply Token).

— **Full Frontal:** A Face Image Type that specifies frontal images with sufficient resolution for human examination as well as reliable computer face recognition. This type of Face Image Type includes the full head with all hair in most cases, as well as neck and shoulders. This image type is suitable for permanent storage of the face information, and it is applicable to portraits for passport, driver license, and “mugshot” images.

— **Token Frontal:** A Face Image Type that specifies frontal images with a specific geometric size and eye positioning based on the width and height of the image. This image type is suitable for minimizing the storage requirements for computer face recognition tasks such as verification while still offering vendor independence and human verification (versus human examination which requires more detail) capabilities.

— **Post-processed Frontal:** Applying digital post-processing to a captured image can modify this image in a way that it is more suitable for automatic face recognition. The Post-processed Frontal Face Image Type is thought as the interchange format for these kinds of facial images.

— **Basic 3D:** The Basic 3D Image Type is the base Image Type of all 3D Face Image Types. All 3D Face Image Types obey normative requirements of this image type.

— **Full Frontal 3D:** The Full Frontal 3D Image Type combines a Full Frontal 2D image with additional 3D information.

— **Token Frontal 3D:** The Token Frontal 3D Image Type combines a Token Frontal 2D image with additional 3D information.

Table 1 shows the relationships between Face Image Types using the notion of inheritance. For example, Frontal inherits properties from Basic, which means that all normative clauses that apply to Basic also apply to Frontal.

Table 1 — Inheritance of Face Image Types

Face Image Type	Inherits from	Normative clauses	Informative clauses
Basic	None	1, 2, 3, 4, 5, 6	A.1
Frontal	Basic	7	A.2
Full Frontal	Frontal	8	A.3
Token Frontal	Frontal	9	A.4
Post-processed Frontal	Frontal	10	

Figure 1 gives a general overview of the scene, photographic, digitization, and format requirements for the face image types specified in this part of ISO/IEC 19794.

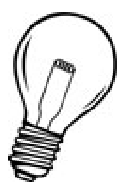
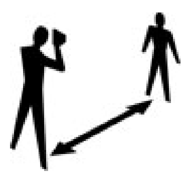
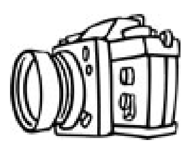
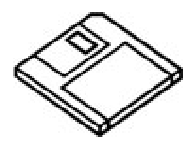
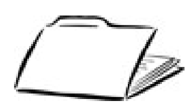
Requirements			
Scene	**Photographic**	**Digital**	**Format**
 Lighting  Image and Subject	 Positioning  Camera Attributes	 Digital Camera  Analogue to Digital  Image Scanning	 Digital Specifications  Record Format and Organization
Clauses: Basic Face None Frontal Face 7.2 Full Frontal Face 8.2	*Clauses:* Basic Face None Frontal Face 7.3 Full Frontal Face 8.3	*Clauses:* Basic Face None Frontal Face 7.4 Full Frontal Face 8.4 Token Face 9.2	*Clauses:* Basic Face 5 6.2 6.3 6.4 Frontal Face 7.5 Full Frontal Face 8.5 Token Face 9.3 Post-processed Frontal Face 10.3

Figure 1 — The types of imaging requirements specified in this part of ISO/IEC 19794. The Basic Face Image Type has no scene, photographic, or digital requirements

2 Conformance

A biometric data record conforms to this part of ISO/IEC 19794 if it satisfies all of the normative requirements related to:

A) Its data structure, data values and the relationships between its data elements, as specified in clauses 5, 6, 7, 8 for the Full Frontal Face Image Type, clauses 5, 6, 7, 9 for the Token Frontal Image Type, and clauses 5, 6, 7, 10 for the Post-processed Frontal Image Type of this part of ISO/IEC 19794, respectively.

B) The relationship between its data values and the input biometric data from which the biometric data record was generated, as specified in clauses 5, 6, 7, 8 for the Full Frontal Face Image Type, clauses 5, 6, 7, 9 for the Token Frontal Image Type, and clauses 5, 6, 7, 10 for the Post-processed Frontal Image Type of this part of ISO/IEC 19794, respectively.

A system that produces biometric data records is conformant to this part of ISO/IEC 19794 if all biometric data records that it outputs conform to this part of ISO/IEC 19794 (as defined above) as claimed in the Implementation Conformance Statement associated with that system. A system does not need to be capable of producing biometric data records that cover all possible aspects of this part of ISO/IEC 19794, but only those that are claimed to be supported by the system in the Implementation Conformance Statement.

A system that uses biometric data records is conformant to this part of ISO/IEC 19794 if it can read, and use for the purpose intended by that system, all biometric data records that conform to this part of ISO/IEC 19794 (as defined above) as claimed in the Implementation Conformance Statement associated with that system. A system does not need to be capable of using biometric data records that cover all possible aspects of this part of ISO/IEC 19794, but only those that are claimed to be supported by the system in an Implementation Conformance Statement.

3 Normative references

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 10918 (all parts), *Information Technology – Digital compression and coding of continuous-tone still images: Requirements and guidelines*

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

ISO/IEC 15444 (all parts), *Information technology — JPEG 2000 image coding system*

ISO/IEC 15948:2003 *PNG, - Portable Network Graphics (PNG): Functional specification*

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

ISO/IEC 29794-1, Information technology — Biometric Sample Quality - Part 1: Framework

4 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 19794-1 and the following apply.

4.1
2D Image
two-dimensional representation that encodes the luminance and/or colour texture of a capture subject in a given lighting environment

4.2
3D Image
image that encodes a surface in a 3D space

4.3
3D Point Map
3D point cloud representing a capture subject, where each surface point is encoded with a triplet, representing the x, y and z value of the point in 3D respectively

4.4
3D Vertex Representation
representation using 3D vertices and triangles between these points for coding of a 3D surface

4.5
Anthropometric Landmark
landmark point on the face used for identification and classification of humans

4.6
Anthropometric Landmark Code
two-part code that defines an Anthropometric Landmark uniquely

4.7
Cartesian Coordinate System
3D orthogonal coordinate system

4.8
chin
the central forward portion of the lower jaw

4.9
colour image
continuous tone image that has more than one channel, each of which is coded with one or multiple bits

4.10
colour space
a way of representing colours of pixels in an image. For instance, RGB, YUV and greyscale are typically used in this document

4.11
continuous tone image
image whose channels have more than one bit per pixel

4.12
crown
top of the head, or (if obscured by hair or headwear), where the top of the head/skull would be if it could be seen

4.13
Cylindrical Coordinate System
three-dimensional polar coordinate system describing a point by the three components radius, azimuth and height

4.14
dots per inch (DPI)
measurement of scanner and printer resolution

4.15
facial image
electronic image-based representation of the portrait of a person

4.16
Face Image Type
a category of facial images that satisfy specific requirements

4.17
Feature Point
reference point(s) in a face image as used by face recognition algorithms, commonly referred to as a landmark

EXAMPLE Position of the eyes.

4.18
fish eye
a type of distortion where central objects of the image erroneously appear closer than those at the edge

4.19
greyscale image
continuous tone image that has only one luminance channel coded e.g. with 8 bit; also referred to as a monochrome or black and white image

4.20
human examination
process of careful human comparison of a face image with a person or another face image to ascertain the identity of the respective person by a detailed examination of facial features and structures

4.21
human verification
to validate the identity of a face image by means of comparison with a person or other face image. One-to-one (1:1) comparison

4.22
JPEG
image compression standard specified as ISO/IEC 10918

NOTE The JPEG baseline standard was published as ISO/IEC 10918-1:1994 and ITU-T Rec. T.81.

4.23
JPEG2000
image compression standard specified as ISO/IEC 15444

NOTE The JPEG2000 baseline standard was published as ISO/IEC 15444-1:2000 and ITU-T Rec. T.800.

4.24
pixel
picture element; element on a two-dimensional array that comprises an image

4.25
PNG format
lossless image compression standard specified in ISO/IEC 15948

4.26
portrait
photograph of a person which includes the full head, with all hair in most cases, as well as neck and top of shoulders

4.27
Range Image
numerical matrix that encodes a surface point in 3D space, where the position encodes the first two coordinates and the value at that position encodes the third coordinate

4.28
red-eye
the red glow from a subject's eye caused by light from flash reflecting from blood vessels behind the retina

4.29
texture
two-dimensional representation of the luminance and/or colour of a capture subject in a given lighting environment

4.30
Texture Projection Matrix
3x4 matrix to transform a 3D surface coordinate from a metric Cartesian Coordinate System to a 2D texture image coordinate, where the transformation makes use of the 3D homogenous coordinates of the 3D point as well as the 2D homogenous coordinates of the 2D point.

NOTE see ref [2] for details

4.31
verification
biometric system function that performs a one-to-one comparison

5 The ISO/IEC 19794-5 Biometric Data Interchange Record Format

5.1 Overview

The ISO/IEC 19794-5 BDIR format specified in this document is a format to store face representations within a biometric data record. Each BDIR shall pertain to a single subject and shall contain at least one or more 2D image and zero or more geometric representations (range images, 3D point maps, 3D vertex representations) of a human face. Depending on the face image type, a 3D representation of a face may be included in addition to the 2D representation. The record structure is depicted in Figure 2, and Figure 3.

Adherence to this format requires compliance to the standards referred to above. In particular, 2D image data will be encoded using JPEG, JPEG2000 or PNG.

When referring to elements of the record format, “field” denotes the minimum element such as Face Image Type and Image Data Type, “block” denotes the group of fields such as Facial Information block or Image Information block, and “record” denotes the data that consists of the General Header and one or more Representations.

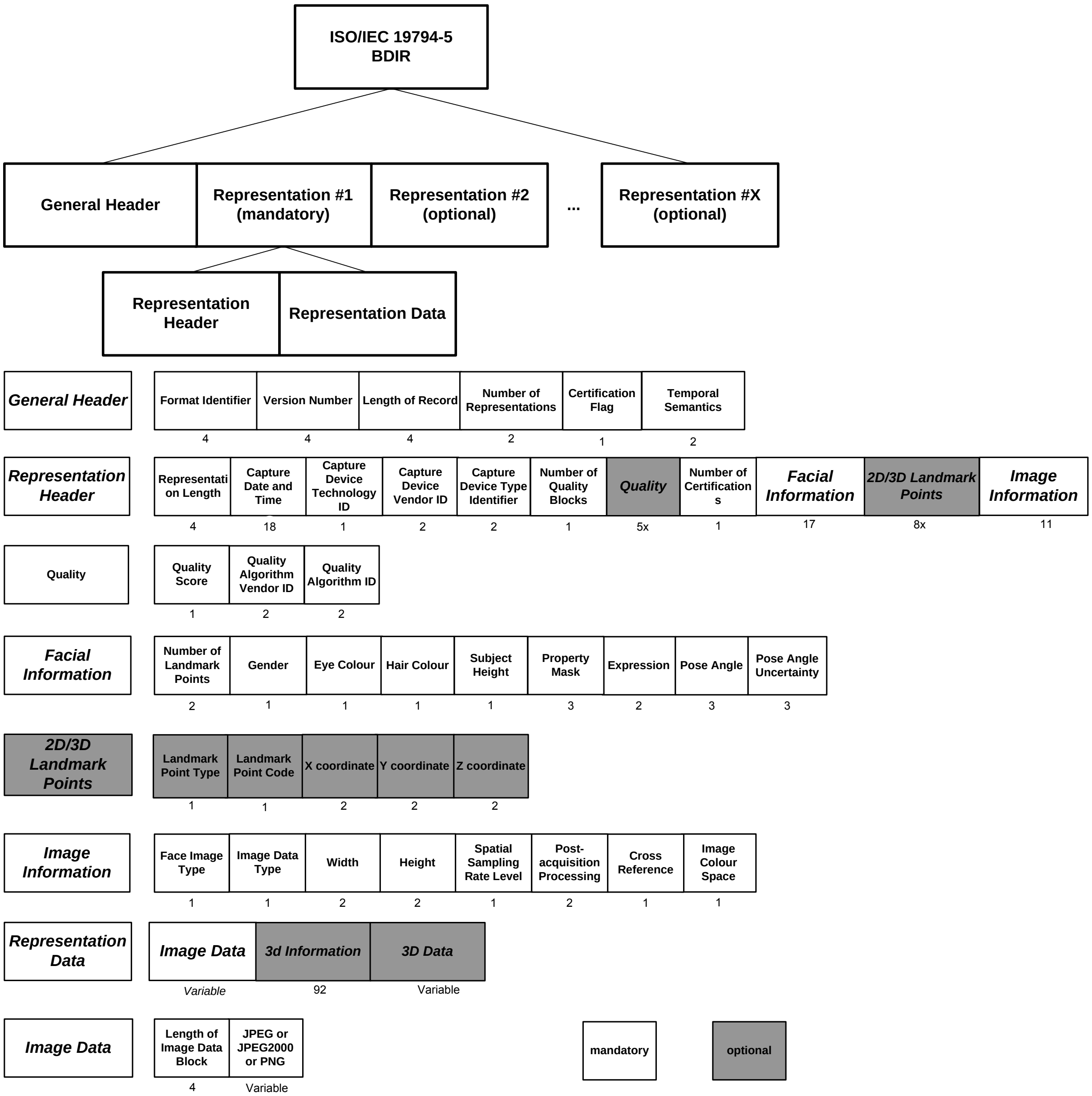


Figure 2 — The ISO/IEC 19794-5 Biometric Data Interchange Record. The length value of each field in bytes is shown below the field. The white boxes indicate fields or blocks that shall be specified, and dark grey boxes indicate optional fields.

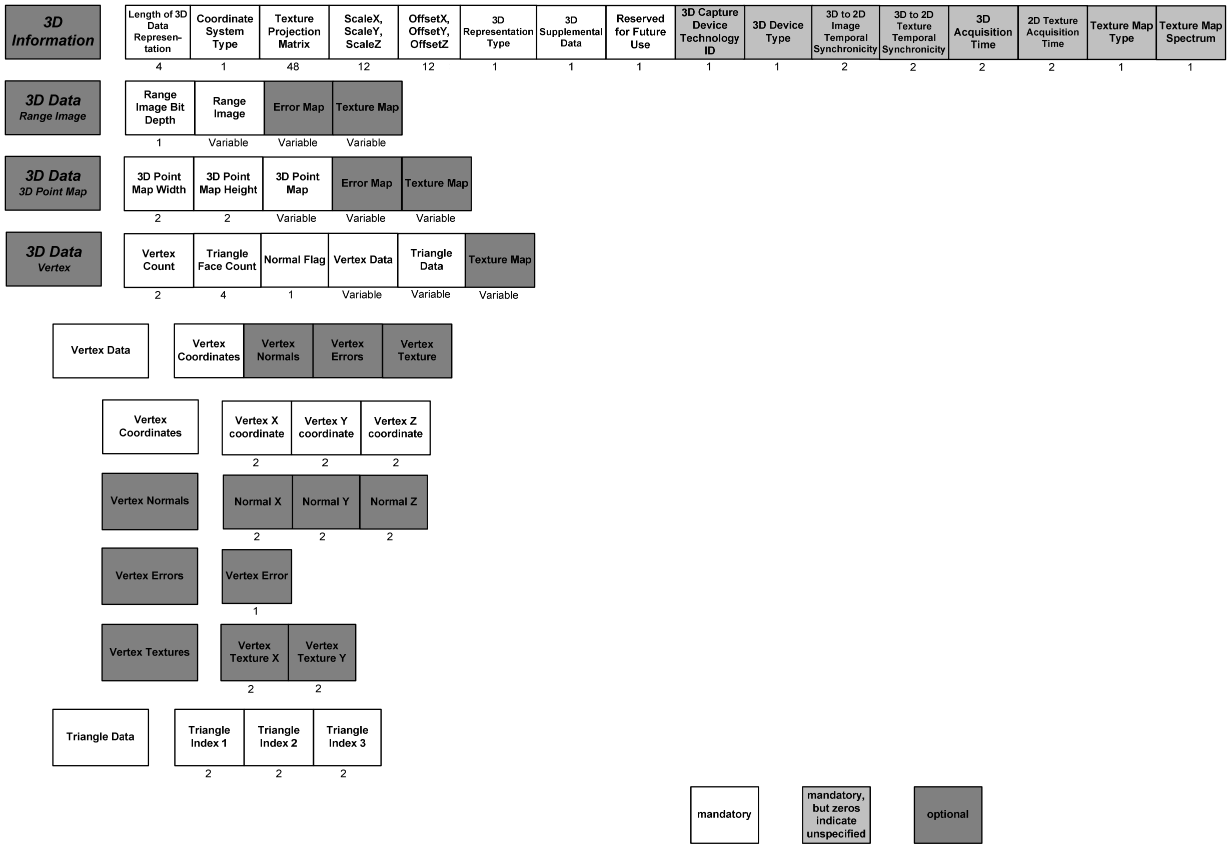


Figure 3 — The 3D Information block and the three possible 3D Data blocks specified in this standard. The length value of each field in bytes is shown below the field. The white boxes indicate fields or blocks that shall be specified, light grey boxes that the fields are mandatory, but an unspecified value is acceptable, and dark grey boxes indicate optional fields.

With the exception of the Format Identifier and the Version Number for the standard, which are null-terminated ASCII character strings, all data is represented in binary format.

There are no record separators or field tags; fields are parsed by byte count.

The organization of the record format is as follows:

- A fixed-length (17 byte) **General Header** containing information about the overall record, including the number of facial images represented and the overall record length in bytes;
- A Representation block for each facial representation. This data consists of a Representation Header and the Representation Data.
- The Representation Header consists of
  - o Multiple (including none) fixed length (5 byte) **Quality** blocks describing the quality of the representation.
  - o A fixed length (21 byte) **Facial Information** block describing discernable characteristics of the subject such as gender.
  - o Multiple (including none) fixed length (8 byte) **Landmark Point** blocks describing Landmark Points in a facial image.
  - o A fixed length (25 byte) **Image Information** block describing digital properties of the image such as Face Image Type and dimensions such as width and height.

- The Representation Data consists of
  - **Image data** consisting of a JPEG, JPEG2000 or PNG encoded data block.
  - For Face Image Types containing 3D information a **3D Information** block (92 byte) describing properties of this data.
  - For Face Image Types containing 3D information the **3D Data** block describing the 3D shape of the face.

Multiple 2D / 3D representations of the same biometric data subject can be described in a single record. This is accomplished by including multiple representation blocks after the General Header block. Representation blocks containing 2D data can be stored together with Representation blocks also containing 3D data. The structure of this embedding is illustrated in Figure 4.

General Header	**Representation #1 (Full Frontal)**	**Representation #2 (Token Frontal)**	**Representation #3 (Full Frontal Range)**

Figure 4 — Embedding multiple 2D / 3D representations in the same record

5.2 Data Conventions

5.2.1 Byte ordering

Within the record format and all well-defined data blocks therein, all multi-byte quantities are stored in Big-Endian format. That is, the more significant bytes of any multi-byte quantity are stored at lower addresses in memory than less significant bytes. For example, the value 1 025 (2 to the 10^{th} power plus one) would be stored as first byte= 00000100 and second byte=00000001.

5.2.2 Numeric values

All numeric values are fixed-length unsigned integer quantities, unless otherwise specified.

5.2.3 Conversion to integer

The conversion of a numeric value to integer is given by rounding down if the fractional portion is less than 0,5 and rounding up if the fractional value is greater than or equal to 0,5.

5.2.4 Unspecified field value

In Figure 2 and Figure 3 the following fields are mandatory, but a value of zero codes that the field is unspecified: Capture Device Technology ID, Capture Device Vendor ID, Capture Device Type Identifier, Gender, Eye Colour, Hair Colour, Subject Height, Expression, Pose Angle, Pose Angle Uncertainty, Image Colour Space, 3D Capture Device Technology ID, 3D Device Type, Texture Map Type, and Texture Map Spectrum.

5.2.5 Unknown field value

A field value labelled by the identifier "Unknown" shall be used to denote that the information encoded by the field cannot be determined by examination of the face image.

5.3 The General Header

5.3.1 Organization

The General Header block consists of seven fields; Format Identifier, Version Number, Length of Record, Number of Representations, Capture Device Vendor ID, Capture Device Type ID and the Temporal Semantics field as shown in Table 2.

Table 2 — The General Header

Field	Size	Valid values	Notes
Format Identifier	4 bytes	0x46414300 (‘F’ ‘A’ ‘C’ 0x0)	Indicates face representation data
Version Number	4 bytes	0x30333000 (‘0’ ’3’ ’0’ 0x0)	“030” in ASCII
Length of Record	4 bytes	69 ≤ Length of Record ≤ 2^{32} - 1	See 5.3.4
Number of Representations	2 byte	1 ≤ Number ≤ 65535	See 5.3.5
Certification Flag	1 byte	0x00 0x01	See 5.3.6
Temporal Semantics	2 bytes	0 ≤ Number ≤ 65535	See 5.3.7 and Table 3

5.3.2 Format Identifier

The format identifier shall be recorded in four bytes. The format identifier shall consist of three characters "FAC" followed by a zero byte as a NULL string terminator..

5.3.3 Version Number

The number for the version of ISO/IEC 19794-5 used for constructing the BDIR shall be placed in four bytes. This version number shall consist of three ASCII numerals followed by a zero byte as a NULL string terminator. The first and second character will represent the major version number and the third character will represent the minor revision number. The Version Number of ISO/IEC 19794-5:2010 shall be 0x30333000; “030” – Version 3 revision 0.

5.3.4 Length of Record

The length (in bytes) of the entire BDIR shall be recorded in four bytes. This count shall be the total length of the BDIR including the general record header and one or more representation records.

5.3.5 Number of Representations

The total number of representation records contained in the BDIR shall be recorded in two bytes. A minimum of one representation is required.

5.3.6 Certification Flag

The one-byte Certification Flag shall indicate whether each Representation Header includes a certification block. A value of 00Hex shall indicate that no representation contains a certification block. A value of 01Hex shall indicate that all representations contain a certification block.

NOTE A certification block that is present may contain 0 certifications (in that case the number-of-certifications field in the certification block has the value 0).

5.3.7 Temporal Semantics

This two byte (2 byte) field shall be assigned according to Table 3. This supports storage of multiple representations: from a single session (e.g. from a photo shoot); from distinct sessions (e.g. from cash dispenser transactions); and from a temporal sequence (e.g. a video sequence of equally time-spaced representations).

Table 3 — Temporal sequence flags and values

Description	Value
One representation is present	0x0000
Two or more representations are present. The temporal relationship between them is unspecified.	0x0001
Two or more representations are present. The representations are taken at irregular intervals during a single session (e.g. images taken by a photographer during a shoot.)	0x0002
Two or more representations are present. The representations are taken at irregular intervals spanning multiple sessions (e.g. a lifetime of passport photo submissions). Unless representation qualities are known, or some other appropriate ordering mechanism is available, the most recent representation should be stored first.	0x0003
The number of milliseconds between sequential representations (e.g. the frames extracted from a video sequence).	4 ≤ Number < 65 534
The representations correspond to a temporal sequence but are taken at regular intervals exceeding 65533 milliseconds.	65 534

NOTE 1 The minimum interval is 4 * 0,001 seconds = 0,004 seconds.

NOTE 2 The maximum interval is 65 533 * 0,001 seconds = 65,533 seconds.

NOTE 3 While a proper video encoding, e.g. ITU-T H264, would offer a more compact encoding of the same sequence, it would add a decoding overhead and is anyway not supported by this standard.

5.4 The Representation Header

5.4.1 Structure

The Representation Header is intended to describe discrete properties of the individual discernable from the image, one is included for each facial representation included in the record. The structure of this block is shown in Figure 2.

The Representation Header consists of the Representation Length, the Capture Date and Time, the Capture Device Technology Identifier, the Capture Device Vendor Identifier. These are followed by the Number of Quality Blocks field and the related number of Quality blocks. Finally the Representation Header contains the Number of Certifications field, the Facial Information block, the optional multiple Landmark Point blocks, and fthe Image Information block.

5.4.2 Representation Length

The (4 byte) Representation Length field denotes the sum of the lengths of the Facial Information block, the 2D/3D Landmark Points block(s), the Image Information block, the Quality block, the Representation data block, the 3D Information block and the 3D Data block.

The minimum value of the Representation Length is 46 bytes (length for the Facial Information block, 21 bytes, plus the Image Information block, 25 bytes) plus the size of the Representation data blocks (in bytes).

5.4.3 Capture Date and Time

The date and time field within a representation header shall be stated in Coordinated Universal Time (UTC). The format given in Table 4 shall also be used for any absolute time values in representation headers. The capture date and time shall be included in the representation header while the biometric sample encoding time is optionally added. Values for the date and time field shall be encoded as BCD.

Table 4 — Format for date and time

#	Part	Name	Allowed values	Example Thursday 17:35:19.7946 December 15, 2005	
				Text Value	**BCD Encoding**
1	YYYY	Four digit Gregorian calendar year	1900 to 9998 9999 = UNASSIGNED	2005	0010 0000 0000 0101
2	MM	Two digit month	01 to 12	12	0001 0010

			99 = UNASSIGNED		
3	DD	Two digit day	01 to 31 99 = UNASSIGNED	15	0001 0101
4	hh	Two digit hour in 24 hour clock	00 to 23 99 = UNASSIGNED	17	0001 0111
5	mm	Two digit minute	00 to 59 99 = UNASSIGNED	35	0011 0101
6	ss	Two digit second	00 to 59 99 = UNASSIGNED	19	0001 1001
7	uuuu	Four digit tenths of milliseconds	0000 to 9998 9999 = UNASSIGNED	7946	0111 1001 0100 0110

5.4.4 Capture Device Technology ID

Capture device technology ID shall be encoded in one byte. This field shall indicate the class of device technology used to acquire the captured biometric sample. A value of 0 indicates “unknown or unspecified”. See Table 5 for the enumerated list of possible values.

EDITOR’s NOTE: Based on the DOC of a GB comment this harmonized text has to be appended with:

Many different types of capture devices work in the visible spectrum or in near infra-red (NIR). To indicate that the capture device operates in NIR the highest bit in the Capture Device Technology ID field shall be set to 1. So, if the representation is a Video frame from a digital video camera operating in the visible spectrum, the resulting value of the Capture Device Technology ID field shall be 0x06, if it operates in NIR, the code shall be 0x86.

Table 5 — Capture Device Technology ID codes

Description	Value
Unknown or Unspecified	0x00
Static photograph from an unknown source	0x01
Static photograph from a digital still-image camera	0x02
Static photograph from a scanner	0x03
Video frame(s) from an unknown source	0x04
Video frame(s) from an analogue video camera	0x05
Video frame(s) from a digital video camera	0x06
Static photograph from a dedicated near infra-red camera	0x07
Reserved by SC37	0x08 – 0x7F
Vendor Specific	0x88 – 0xFF

5.4.5 Capture Device Vendor Identifier

The (2 byte) Capture Device Vendor Identifier shall identify the biometric organisation that owns the product that created the BDIR. The capture device algorithm vendor identifier shall be encoded in two bytes carrying a CBEFF biometric organization identifier (registered by IBIA or other approved registration authority). A value of all zeros shall indicate that the capture device vendor is unreported.

5.4.6 Capture Device Type Identifier

The (2 byte) Capture Device Type Identifier shall identify the product type that created the BDIR. It shall be assigned by the registered product owner or other approved registration authority. A value of all zeros shall indicate that the capture device type is unreported.

5.4.7 Number of Quality Blocks

This field is followed by the number of 5 byte Quality blocks reflected by its value. A value of zero (0) means that no attempt was made to assign a quality score. In this case, no Quality blocks are present.

5.4.8 Quality Score

The (1 byte) Quality Score, as defined in ISO/IEC 29794-1, shall be a quantitative expression of the predicted verification performance of the biometric sample. Valid values for Quality Score are integers between 0 and 100, where higher values indicate better quality. A value of 255 is to handle a special case. An entry of 255 shall indicate a failed attempt to calculate a quality score. This value of Quality Score is harmonized with ISO/IEC 19784-1, where 255 is -1.

NOTE BioAPI, unlike ISO/IEC 19794 uses signed integers

5.4.9 Quality Algorithm Vendor ID

To enable the recipient of the quality score to differentiate between quality scores generated by different algorithms, the provider of quality scores shall be uniquely identified by this two-byte field. This is registered with the IBIA or other approved registration authority.

5.4.10 Quality Algorithm ID

The (2 byte) Quality Algorithm ID specifies an integer product code assigned by the vendor of the quality algorithm. It indicates which of the vendor's algorithms (and version) was used in the calculation of the quality score and should be within the range 1 to 65 535.

NOTE Multiple quality scores calculated by the same algorithm (same vendor ID and algorithm ID) shall not be present in a single representation.

Table 6 summarizes the quality field. All values are fixed-length unsigned integer quantities represented in Big-Endian format.

Table 6 — Structure of Quality blocks

<table>
<tr><th colspan="2">Description</th><th>Length</th><th>Valid values</th><th>Note</th></tr>
<tr><td colspan="2">Number of Quality Blocks</td><td>1 byte</td><td>[0,255]</td><td>This field is followed by the number of 5-byte Quality Blocks reflected by its value.
A value of zero (0) means that no attempt was made to assign a quality score. In this case, no Quality Blocks are present.</td></tr>
<tr><td rowspan="3">Quality Block</td><td>Quality Score</td><td>1 byte</td><td>[0,100]
255</td><td>0: lowest
100: highest
255: failed attempt to assign a quality score</td></tr>
<tr><td>Quality algorithm vendor ID</td><td>2 bytes</td><td>[1,65535]</td><td>Quality Algorithm Vendor ID shall be registered with IBIA or other approved registration authority as a CBEFF biometric organization. Refer to CBEFF vendor ID registry procedures in ISO/IEC 19785-2.</td></tr>
<tr><td>Quality algorithm ID</td><td>2 bytes</td><td>[1,65535]</td><td>Quality Algorithm ID may be optionally registered with IBIA or other approved registration authority as a CBEFF Product Code. Refer to CBEFF product registry procedures in ISO/IEC 19785-2.</td></tr>
</table>

5.4.11 Number of Certifications

As this part of ISO/IEC 19794 does not support certifications the (1 byte) Number of Certifications field shall be zero.

5.5 The Facial Information Block

5.5.1 Number of Landmark Points

The (2 byte) Number of Landmark Points field shall be the number of Landmark Point blocks that follow the Facial Information block. The Landmark Point block is defined in clause 5.6.

5.5.2 Gender

The (1 byte) Gender field shall represent the gender of the subject according to Table 7.

Table 7 — Gender codes

Description	Value
Unspecified	0x00
Male	0x01
Female	0x02
Unknown	0xFF

5.5.3 Eye Colour

The (1 byte) Eye Colour field shall represent the colour of irises of the eyes according to Table 8. If the eyes are different colours, then the right eye colour is to be encoded.

Table 8 — Eye Colour codes

Description	Value
Unspecified	0x00
Black	0x01
Blue	0x02
Brown	0x03
Gray	0x04
Green	0x05
Multi-Coloured	0x06
Pink[1]	0x07
Reserved by SC37	0x08 – 0xFE
Other or Unknown (e.g. cannot be determined from image, monochrome image)	0xFF

5.5.4 Hair Colour

The (1 byte) Hair Colour field shall represent the colour of the hair according to Table 9.

Table 9 — Hair Colour codes

Description	Value
Unspecified	0x00
Bald	0x01
Black	0x02
Blonde	0x03
Brown	0x04
Gray	0x05
White	0x06
Red	0x07
Reserved by SC37	0x08 – 0xFE
Unknown or Other	0xFF

5.5.5 Subject Height

The (1 byte) Subject Height field shall represent the height of the subject according to Table 10.

[1] Very rare, associated to albinism. Very small minority of individuals world wide are affected.

Table 10 — Subject Height codes

Description	Value
Unspecified	0x00
Subject Height in cm	0x01-0xFF

5.5.6 Property Mask

The (3 byte) Property Mask is a bit mask of 3 bytes and each bit of the mask position listed in Table 11 shall be set to 1 if the corresponding property is present, and set to 0 if absent. The mask position starts from 0 at the lowest bit. The lowest bit set to 0 shall indicate that properties are not specified; the lowest bit set to 1 shall indicate that all listed properties have been considered and that a zero value of any property bit indicates an absence of that property.

Table 11 — Property flags

Description	Mask Position
Properties are specified	0
Glasses	1
Moustache	2
Beard	3
Teeth visible	4
Pupil or iris not visible (e.g. either or both eyes closed or half closed)	5
Mouth open	6
Left Eye Patch	7
Right Eye Patch	8
Dark Glasses (medical)	9
Head coverings present	10
Feature Distorting Medical Condition (which could impact Feature Point detection)	11
Reserved for future definition by SC 37	12 – 23

Note that a “Pupil or iris not visible” flag set to “1” will indicate non-compliance with the Frontal, Full Frontal, and Token image types.

5.5.7 Expression

The (2 byte) Expression field shall represent the expression of the face according to Table 12.

Table 12 — Expression codes

Description	High Byte	Low Byte
Unspecified	0x00	0x00
Neutral (non-smiling) with both eyes open and mouth closed	0x00	0x01
Smile	0x00	0x02
Raised eyebrows	0x00	0x03
Eyes looking away from the camera	0x00	0x04
Squinting	0x00	0x05
Frowning	0x00	0x06
Reserved for future definition by SC 37	0x00 0x01 – 0x7F	0x07 – 0xFF 0x00 – 0xFF
Reserved for vendor definition	0x80 – 0xFF	0x00 – 0xFF

5.5.8 Pose Angle

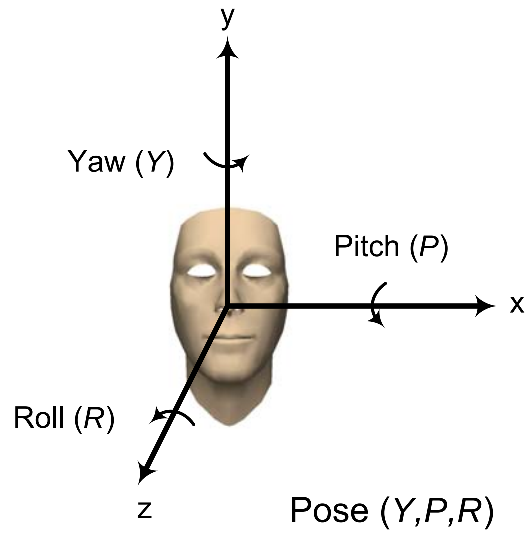


Figure 5 — The definition of pose angles is with respect to the frontal view of the subject

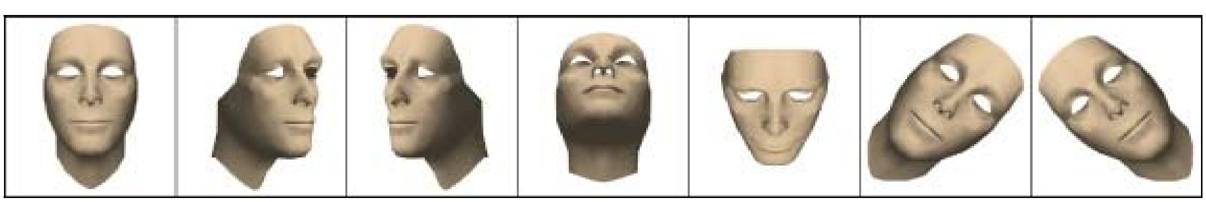
The (3 multi-byte) Pose Angle field (B_Y, B_P, B_R) shall represent the estimate or measure pose of the subject in the image. Each byte in the field respectively represents pose angles of yaw, pitch and roll in that order. The pose angle is given by Tait-Bryan angles.

- Yaw angle: rotation about the vertical (y) axis.
- Pitch angle: rotation about the horizontal side-to-side (x) horizontal axis.
- Roll angle: rotation about the horizontal back to front (z) axis.

The angles are defined relative to the frontal pose of the subject, which has angles (0,0,0) as shown in Figure 5. The frontal pose is defined by the Frankfurt Horizon FH (see Annex D) as the xz plane and the vertical symmetry plane as the yz plane with the z axis oriented in the direction of the face sight. The examples are shown in Figure 6.

As order of the successive rotation around the different axes does matter, the encoded rotation angle shall correspond to an order of execution starting from the frontal view. This order shall be given by Roll (about the front axis), then Pitch (about the horizontal axis) and finally Yaw (about the vertical axis). The (first executed) Roll transformation will therefore always be in the image (x,y) plane.

From the point of view of executing a transformation from the observed view to a frontal view, the transformation order will therefore be Yaw, Pitch, and then Roll. Note however that the encoded angle is from the frontal view to the observed view.



A	b	C	D	e	f	g

Figure 6 — Examples of pose angles and their encodings. The pose angles (Y, P, R) of figures a – g are given by (0, 0, 0), (+45, 0, 0), (-45, 0, 0), (0, -45, 0), (0, +45, 0), (0, 0, -45), and (0, 0, +45), respectively. The pose angle encodings, (B_Y, B_P, B_R) as defined in clauses 5.5.8.1 - 5.5.8.3 are subsequently given by (1, 1, 1), (23, 1, 1), (158, 1, 1), (1, 158, 1), (1, 23, 1), (1, 1, 158), and (1, 1, 23), respectively

5.5.8.1 Pose Angle – Yaw

The yaw angle Y is the rotation in degrees about the y-axis (vertical axis) shown in Figure 5. Frontal poses have a yaw angle of 0 degrees. Positive angles represent faces looking to their left (a counter-clockwise rotation around the y-axis).

The encoding B_Y of the yaw angle Y shall be in degrees as a byte (1 byte) with values from –180 to 180 modulo 2.

- If $180 \geq Y \geq 0$ then $B_Y = Y/2+1$. The remainder is discarded.
- If $-180 \leq Y < 0$ then $B_Y = 181+Y/2$. The remainder is discarded.

The maximum value of B_Y is 181. If the pose angle is not specified, the value of B_Y shall be 0.

5.5.8.2 Pose Angle – Pitch

The pitch angle P is the rotation in degrees about the x-axis (horizontal axis) shown in Figure 5. Frontal poses have a pitch angle of 0 degrees. Positive angles represent faces looking down (a counter-clockwise rotation around the x-axis).

The encoding B_P of the pitch angle P shall be in degrees as a byte (1 byte) with values from –180 to 180 modulo 2.

- If $180 \geq P \geq 0$ then $B_P = P/2+1$. The remainder is discarded.
- If $-180 \leq P < 0$ then $B_P = 181+P/2$. The remainder is discarded.

The maximum value of B_P is 181. If the pitch angle is not specified, the value of B_P shall be 0.

Note that the definition of zero degree pitch is not obvious.

5.5.8.3 Pose Angle – Roll

The roll angle R is the rotation in degrees about the z-axis (the horizontal axis from front to back) shown in Figure 5. Frontal poses have a roll angle of 0 degrees. Positive angles represent faces tilted toward their right shoulder (counter-clockwise rotation around the z-axis). A roll angle of 0 degrees denotes that the left and right eye centres have identical Y coordinates.

The encoding B_R of the roll angle R shall be in degrees as a byte (1 byte) with values from –180 to 180 modulo 2.

- If $180 \geq R \geq 0$ then $B_R = R/2+1$. The remainder is discarded.
- If $-180 \leq R < 0$ then $B_R = 181+R/2$. The remainder is discarded.

The maximum value of B_R is 181. If the roll angle is not specified, the value of B_R shall be 0.

5.5.9 Pose Angle Uncertainty

The (3 multi-byte) Pose Angle Uncertainty (U_Y, U_P, U_R) represents the expected degree of uncertainty of the pose angle yaw, pitch, and roll. Each byte in the field respectively represents the uncertainty of yaw, pitch and roll in that order. The uncertainty is allowed to represent experimental uncertainty specified by each vendor.

The encoding of Pose Angle Uncertainty is given by three bytes (U_Y, U_P, U_R) where each byte U_k in the field (k=Y,P,R) represents 1 degree of uncertainty with minimum and maximum values of 1 and 181 where U_k=(uncertainty+1). The more uncertain, the value of the uncertainty U_k shall become larger. If the uncertainty is not specified, then the values of U_Y, U_P and U_R shall be set to zero (0).

5.6 The Landmark Point Block

5.6.1 Structure

The optional (8 byte) Landmark Point block specifies the type, code and position of a Landmark Point in the facial image. The number of Landmark Point blocks shall be specified in the Number of Landmark Points field of the Facial Information block. The structure of this block is shown in Table 14.

Landmark Points can be specified as MPEG-4 Feature Points as given by Annex C of ISO/IEC 14496-2 or anthropometric landmarks in two or three dimensions. The description of the anthropometric landmarks [10] and their relation with the set of MPEG4 feature points is discussed in clause 5.6.6 of this document.

5.6.2 Landmark Point Type

The (1 byte) Landmark Point Type field represents the type of the Landmark Point stored in the Landmark Point block. This field shall be set to 0x01 to denote that landmark point is an MPEG4 Feature Point as given by Annex C of ISO/IEC 14496-2 and is represented by the 2D image coordinates. The field shall be set to 0x02 to denote that the landmark point is an Anthropometric 2D landmark and is represented by the 2D image coordinates. Finally, the field shall be set to 0x03 to denote that the Landmark Point is an anthropometric 3D landmark and is represented by its 3D coordinates. All other field values are reserved by SC 37 for future definition of Landmark Point Types.

Table 13 — The Landmark Point Type

Description	Value	Comment
MPEG4 Feature	0x01	The horizontal and vertical position of the Landmark Point are measured in pixels with values from 0 to Width-1 and Height-1, respectively. The Z coordinate field is ignored.
Anthropometric 2D landmark	0x02	The Landmark Point is considered as an anthropometric landmark point in the 2-D image and its coordinates are measured in pixels with values from 0 to Width-1 and Height-1, respectively. The Z coordinate field is ignored.
Anthropometric 3D landmark	0x03	X coordinate, Y coordinate and Z coordinate are interpreted as 2 byte values with fixed precision of 0,02 mm ranging from -655,34 mm to 655,34 mm. The Landmark Point is considered as a 3D point in the Cartesian coordinate system. EXAMPLE: The value of 10001 corresponds to: 655,34mm + 10001 x 0,02mm = -455,32mm.
Reserved by SC37	0x04-0xFF	Reserved for future use.

5.6.3 Landmark Point Code

The (1 byte) Landmark Point Code field shall specify the Landmark Point that is stored in the Landmark Point block.

For the Landmark Point Type 0x01 the codes of the Landmark Points in clause 5.6.4, taken from Annex C of ISO/IEC 14496-2 and defined as MPEG4 Feature Points, or the additional eye and nostril Landmark Points in clause 5.6.5 shall be stored in this block.

If the Landmark Point Type is 0x02 or 0x03, i.e. anthropometric 2D landmark or anthropometric 3D landmark, the codes of the Landmark Points defined in 5.6.6 shall be stored in this block.

The horizontal and vertical positions of Landmark Points are either texture image coordinates or in the Cartesian coordinate system (see clause 5.10.2.2).

Table 14 — The Landmark Point block

Field	Size	Value	Notes
Landmark Point Type	1 byte	See clause 5.6.2	Denotes the type of the Landmark Point.
Landmark Point Code	1 byte	See clause 5.6.3	Denotes the Landmark Point, e.g. the left eye.

X coordinate, Y coordinate	2 bytes	See clause 5.6.2, Table 13	Denotes the coordinate of the landmark point. For Landmark Point Types 0x01 and 0x02 this coordinate denotes the relevant pixel count from upper left pixel starting at 0. For Landmark Point Type 0x03 the value codes the coordinate of a point in 3D.
Z coordinate	2 bytes	See clause 5.6.2, Table 13	Denotes the Z coordi- nate of the landmark point. For Landmark Point Type 0x01 and Type 0x02 this field is ignored. For Landmark Point Type 0x03 the value codes the Z coordinate of a point in 3D.

5.6.4 MPEG4 Landmark Points

The normative Figure 7 denotes the Landmark Point codes associated with Feature Points as given by Annex C of ISO/IEC 14496-2. Each Landmark Point Code is represented by a notation A.B using a major (A) and a minor (B) value. The encoding of the Landmark Point Code is given by the (1 byte) value of A*16 + B.

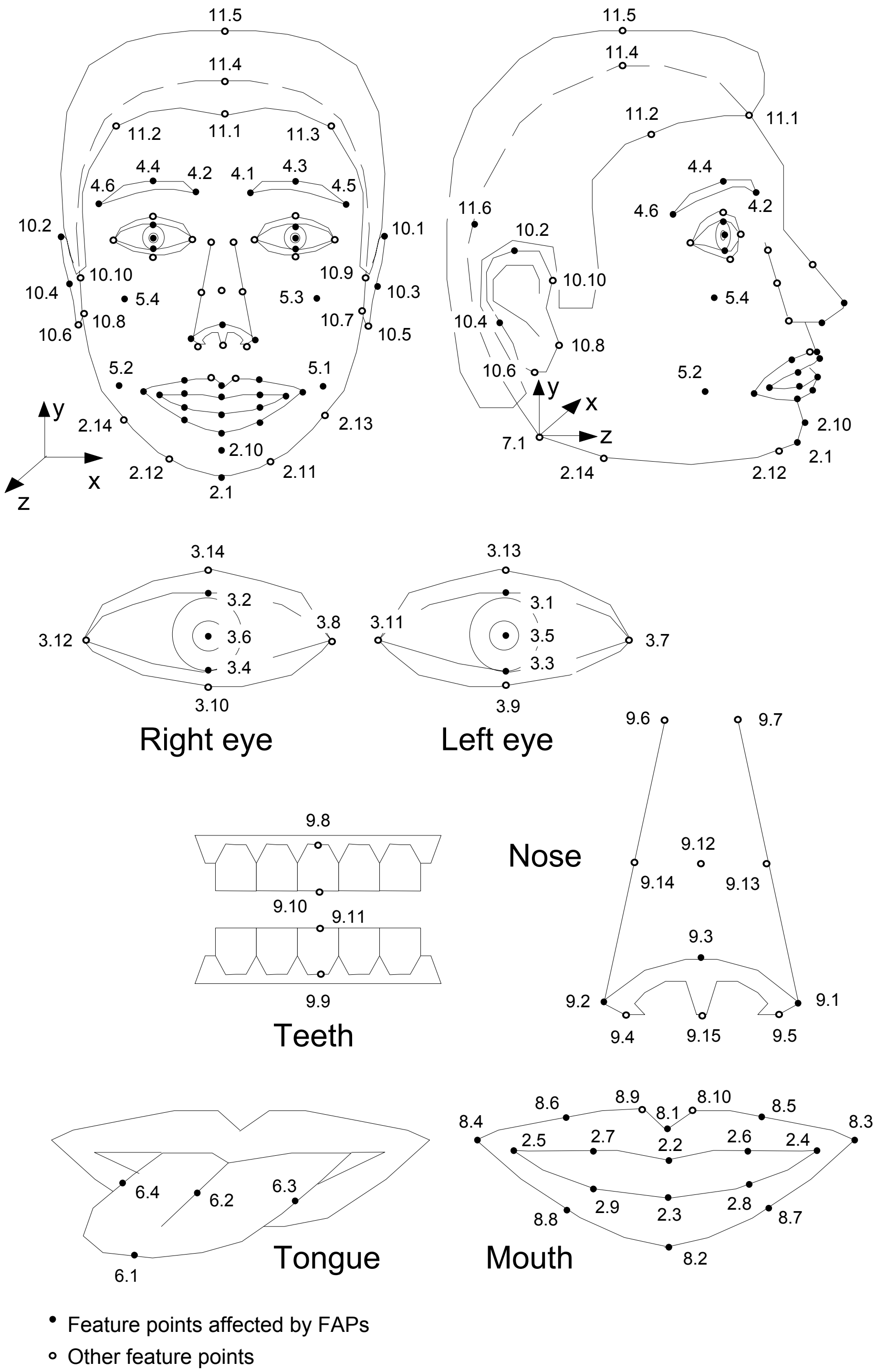


Figure 7 — Feature points affected by Face Animation Parameters specified in ISO/IEC 14496-2.

5.6.5 Eye and nostril centre Landmark Points

The eye centre Landmark Points 12.1 (left) and 12.2 (right) are defined to be the horizontal and vertical midpoints of the eye corners (3.7, 3.11) and (3.8, 3.12) respectively. The left nostril centre Landmark Point 12.3 is defined to be the midpoint of the nose Landmark Points (9.1, 9.15) in the horizontal direction and (9.3, 9.15) in the vertical direction. Similarly, the right nostril centre Landmark Point 12.4 is defined to be the midpoint of the nose Landmark Points (9.2, 9.15) in the horizontal direction and (9.3, 9.15) in the vertical direction. Both the eye centre and nostril centre Landmark Points are shown in Figure 8 and values given in Table 15 — .

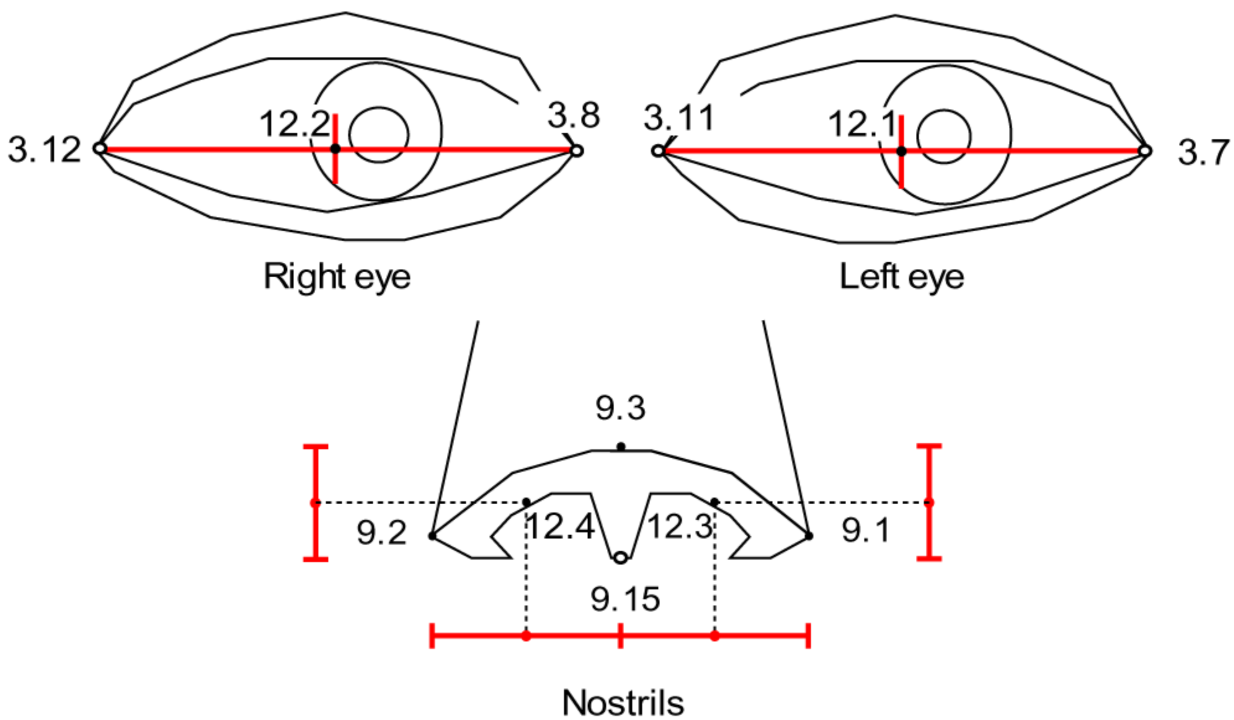


Figure 8 — The eye and nostril centre Landmark Points are defined by midpoints of MPEG4 Landmark Points

Table 15 — Eye and nostril centre Landmark Point Codes

Centre Landmark Point	Midpoint of Landmark Points		Landmark Point Code
Left Eye	3.7, 3.11		12.1
Right Eye	3.8, 3.12		12.2
Left Nostril	Horizontal	Vertical	12.3
	9.1, 9.15	9.3,9.15	
Right Nostril	Horizontal	Vertical	12.4
	9.2, 9.15	9.3,9.15	

5.6.6 Anthropometric Landmarks

Anthropometric landmarks extend the MPEG4 feature model with new points that are used in forensics and anthropology for person identification via two facial images or image and skull over a long time. They also allow specification of points that are in use by criminal experts and anthropologists [10].

Figure 9 and Table 16 show the definition of the anthropometric landmarks. The set of points represents the craniofacial landmark points of the head and face. The latter are used in forensics for “Face to face” and “Skull to face” identification. Some of these points have MPEG 4 counterparts, others not.

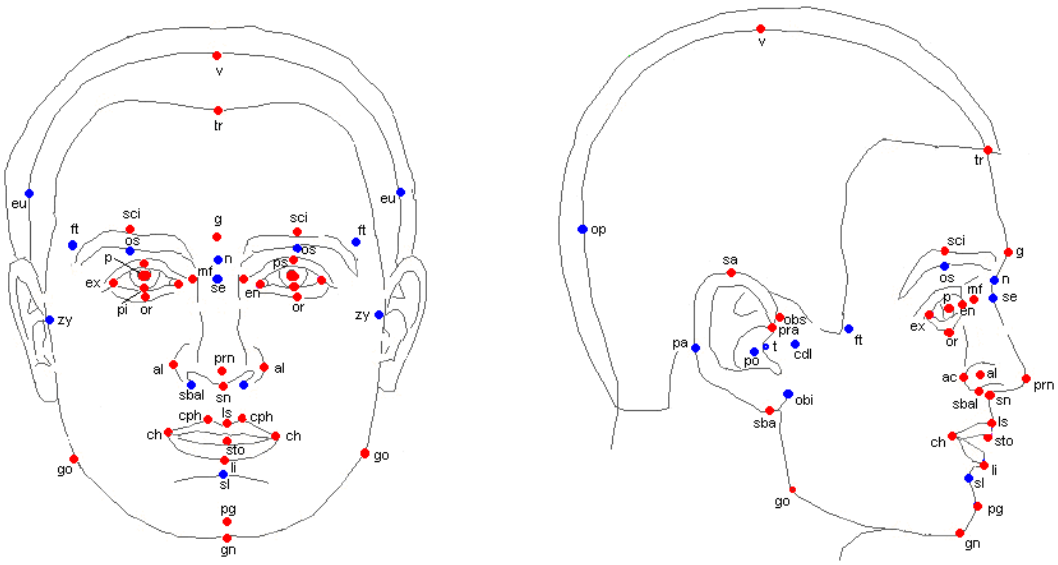


Figure 9: Anthropometric landmarks with (red) and without (blue) MPEG4 counterparts.

Definitions for these points are presented in the Table 16.

Table 16 — Definitions of the anthropometric landmarks

Point ID	Point Code	MPEG4	Anthropometric point name	How to point
V	1.1	11.4	vertex	The highest point of head when the head is oriented in Frankfurt Horizon. Refer to Annex D for the definition of the Frankfurt Horizon.
G	1.2		glabella	The most prominent middle point between the eyebrows
op	1.3		opisthocranion	Situated in the occipital region of the head is most distant from the glabella
eu	1.5, 1.6		eurion	The most prominent lateral point on each side of the skull in the area of the parietal and temporal bones
ft	1.7, 1.8		frontotemporale	The point on each side of the forehead, laterally from the elevation of the linea temporalis
tr	1.9	11.1	trichion	The point on the hairline in the midline of the forehead
zy	2.1, 2.2		zygion	The most lateral point of each of the zygomatic
go	2.3, 2.4	2.15, 2.16	gonion	The most lateral point on the mandibural angle close to the bony gonion
sl	2.5		sublabiale	Determines the lower border of the lower lip or the upper border of the chin
pg	2.6	2.10	pogonion	The most anterior midpoint of the chin, located on the skin surface in the front of the identical bony landmark of the mandible
gn	2.7	2.1	menton (or gnathion)	The lowest median landmark on the lower border of the mandible
cdl	2.9, 2.10		condylion laterale	The most lateral point on the surface of the condyle of the mandible
En	3.1, 3.2	3.11, 3.8	endocanthion	The point at the inner commissure of the eye fissure
Ex	3.3, 3.4	3.7, 3.12	exocanthion (or ectocanthion)	The point at the outer commissure of the eye fissure

P	3.5, 3.6	3.5, 3.6	center point of pupil	Is determined when the head is in the rest position and the eye is looking straight forward
Or	3.7, 3.8	3.9, 3.10	orbitale	The lowest point on the lower margin of each orbit
Ps	3.9, 3.10	3.1, 3.2	palpebrale superius	The highest point in the midportion of the free margin of each upper eyelid
Pi	3.11, 3.12	3.3, 3.4	palpebrale inferius	The lowest point in the midportion of the free margin of each lower eyelid
Os	4.1, 4.2		orbitale superius	The highest point on the lower border of the eyebrow
Sci	4.3, 4.4	4.3, 4.4	superciliare	The highest point on the upper border in the midportion of each eyebrow
N	5.1		nasion	The point in the middle of both the nasal root and nasofrontal suture
Se	5.2		sellion (or subnasion)	Is the deepest landmark located on the bottom of the nasofrontal angle
Al	5.3, 5.4	9.1, 9.2	alare	The most lateral point on each alar contour
Prn	5.6	9.3	pronasale	The most protruded point of the apex nasi
Sn	5.7	9.15	subnasale	The midpoint of the angle at the columella base where the lower border of the nasal septum and the surface of the upper lip meet
Sbal	5.9, 5.10		subalare	The point at the lower limit of each alar base, where the alar base disappears into the skin of the upper lip
Ac	5.11, 5.12	9.1, 9.2	alar curvature (or alar crest) point	The most lateral point in the curved base line of each ala
mf	5.13, 5.14	9.6, 9.7	maxillofrontale	The base of the nasal root medially from each endocanthion
cph	6.1, 6.2	8.9, 8.10	christa philtri landmark	The point on each elevated margin of the philtrum just above the vermilion line
ls	6.3	8.1	labiale (or labrale) superius	The midpoint of the upper vermillion line
li	6.4	8.2	labiale (or labrale) inferius	The midpoint of the lower vermillion line
ch	6.5, 6.6	8.3, 8.4	cheilion	The point located at each labial commissure
sto	6.7		stomion	The imaginary point at the crossing of the vertical facial midline and the horizontal labial fissure between gently closed lips, with teeth shut in the natural position
sa	7.1, 7.2	10.1, 10.2	superaurale	The highest point of the free margin of the auricle
sba	7.3, 7.4	10.5, 10.6	subaurale	The lowest point of the free margin of the ear lobe
pra	7.5, 7.6	10.9, 10.10	preaurale	The most anterior point on the ear, located just in front of the helix attachment to the head
pa	7.7, 7.8		postaurale	The most posterior point on the free margin of the ear
obs	7.9, 7.10	10.3, 10.4	otobasion superious	The point of attachment of the helix in the temporal region
obi	7.11, 7.12		otobasion infrious	The point of attachment of the ear lobe to the cheek
po	7.13, 7.14		porion (soft)	The highest point of the upper margin of the cutaneous auditory meatus
t	8.1, 8.2		tragion	The notch on the upper margin of the tragus

The anthropometric landmark code has the format: A.B. A specifies the global landmark of the face to which this landmark belongs such as nose, mouth, etc. B specifies the particular point. In case a Landmark Point has two symmetrical entities (left and right) the right entity always has a greater and even minor code value. Hence, all Landmark Points from the left part of the face have odd minor codes, and from the right part – even minor codes. Both A and B are in the range from 1 to 15. Hence, the code A*16 + B is written to the 1 byte Landmark Point Code field.

5.6.7 Anthropometric 3D landmark

The error of an anthropometric 3D landmark point location should be no greater than 3 mm. The point shall withstand from the nearest point on the surface no further than 3 mm. The point on the surface is a vertex, or a point on an edge, or a point on a face of the surface.

5.6.8 Z coordinate

This field is not used if the Landmark Point Type is equal to MPEG4 feature or anthropometric 2D landmark. In case the Landmark Point Type equals anthropometric 3D landmark this field along with the horizontal and vertical positions denotes the coordinates of the landmark point in the 3D Cartesian coordinate system. The metric coordinates of 3D landmarks shall be obtained by multiplying the X, Y, and Z coordinates by a fixed scale of 0,02 mm. Note, that the Landmark Point Type field codes the type of the Landmark Point and determines the interpretation of the Z coordinate.

5.7 The Image Information Block

5.7.1 Data Structure

The (25 byte) Image Information block is intended to describe digital properties of the facial image, one is included for each facial image included in the record. The structure of this block is shown in Figure 2. One Representation data block shall follow this block.

5.7.2 Face Image Type

The Face Image Type field shall represent the type of the facial image stored in the Image Information block and, if applicable, the 3D Data block according to Table 17.

Note that all Frontal Image Types are either Full Frontal, Token Frontal, Post-processed Frontal or one of the respective 3D Full Frontal or Token Frontal Image Types. Therefore a separate Frontal Value is not required.

Table 17 — Face Image Type codes

Description	Value
Basic	0x00
Full Frontal	0x01
Token Frontal	0x02
Post-processed Frontal	0x03
Reserved by SC37	0x04 – 0x7F
Basic 3D	0x80
Full Frontal 3D	0x81
Token Frontal 3D	0x82
Reserved by SC 37	0x83 - 0xFF

The Basic Face Image Type is defined in clause 6. The Frontal, Frontal/Full, Frontal/Token, Frontal/Post-processed are defined in clauses 7, 8, 9, and 10, respectively. Face Image Types use the notion of inheritance. For example, the Frontal Face Image Type inherits all of the requirements of the Basic Face Image Type - the Frontal Face Image type obeys all normative requirements of the Basic Face Image Type. The inheritance structure of currently defined image types is shown in Figure 10.

If a 2D record that is compliant to the Basic, Full Frontal or Token Frontal requirements, respectively, contains 3D data, this is indicated by the highest bit of the Face Image Type set to one, resulting in the Face Image Type codes 0x80 to 0x82.

Normative requirements for the Basic, Frontal, Full Frontal, Token Frontal and Post-processed Face Image Types are given in clauses 6, 7, 8, 9, and 10 respectively. The 3D Image types are defined in clauses 11, 12, and 13, respectively.

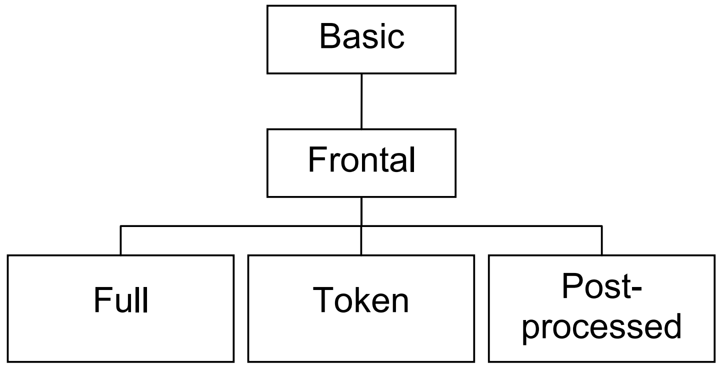


Figure 10 — 2D Face Image Types and their inheritance map

5.7.3 Image Data Type

The (1 byte) Image Data Type field denotes the encoding type of the Image Data block. Either JPEG (ISO/IEC 10918-1 and ITU-T Rec. T.81, [18]) or JPEG2000 (ISO/IEC 15444-1) or PNG (ISO/IEC 15948:2003) shall be specified.

For lossless compression PNG or JPEG2000 lossless shall be used. For lossless representation of images using more than 8 bits per channel PNG or JPEG2000 lossless shall be used. For lossy representation of images using more than eight bit per channel JPEG2000 shall be used. Note that an “Unspecified” Value cannot be encoded.

Table 18 — Image Data Type codes

Description	Value
JPEG	0x00
JPEG2000 lossy	0x01
JPEG 2000 lossless	0x02
PNG	0x03
Reserved by SC 37	0x04 – 0xFF

5.7.4 Width

The (2 byte) Width field shall specify the number of pixels in the horizontal direction.

5.7.5 Height

The (2 byte) Height field shall specify the number of pixels in the vertical direction.

5.7.6 Spatial Sampling Rate Level

For specific application domains different minimal spatial sampling rates of the interchange data may be required. For example using higher spatial sampling rate images allow for specific human as well as machine inspection methods that depend on the analysis of very small details. The (1 byte) Spatial Sampling Rate Level field allows specifying such requirements. The Width of Head CC is defined in Figure 14.

Table 19 — Spatial Sampling Rate Level codes

Width of Head, CC	Spatial Sampling Rate Level
CC ≤ 180	0x00
180 <CC ≤ 240	0x01

240 < CC ≤ 300	0x02
300 < CC ≤ 370	0x03
370 < CC ≤ 480	0x04
480 < CC ≤ 610	0x05
610 < CC ≤ 750	0x06
750 < CC	0x07
Reserved by SC 37	0x08-0xFF

NOTE Interocular distance in pixels will be approximately half of the head widths.

5.7.7 Post-acquisition Processing

While the alteration of face image data is discouraged, there are cases when no alternative may exist:

- legacy database of ¾ frontal face images must be rotated to full frontal for matching,
- from a frontal image artificial non-frontal facial images at predetermined non-frontal poses are automatically generated (multi-view images) using an implicit head model or similar. These images can be beneficial during the comparison process or a manual review process as they show a more similar pose than the original frontal image,
- a single image is to be age progressed and used for verification of a passport holder,
- a short video stream is super-resolved to a single face image for matching against a watchlist.

The (2 byte) Post-acquisition Processing bit field allows the specification of the kind of post processing that has been applied to the original captured image. Each bit of the mask position listed in Table 20 shall be set to 1 if the corresponding processing has been applied, and set to 0 if not. The mask position starts from 0 at the lowest bit. All bits set to zero indicates that no post-acquisition processing has been applied at all.

On the one hand a captured image typically needs some post-processing so that the resulting representatipn conforms to the clauses of this standard, especially for the frontal image type. On the other hand these processing steps should be minimal and not distort the characteristics of the original image. The right column in Table 20 clearly states, what post-acquisition processes can be applied without having to store the resulting representation in the post-processed image type as defined in clause 10.

Table 20 — Post acquisition Processing

Post-acquisition Processing	Mask Position	Use of post-processed image type mandatory
Rotated (in-plane)	0	No
Cropped	1	No
Downsampled	2	No
White balance adjusted	3	No
Multiply compressed	4	Yes
Interpolated (upsample)	5	Yes
Contrast stretched	6	Yes
Pose corrected	7	Yes
Multi View Image	8	Yes
Age progressed	9	Yes
Super-resolution processed	10	Yes
Reserved for future use	11-15	

5.7.8 Cross Reference

The (1 byte) Cross Reference Data Type field denotes inter-dependencies when multiple representations are stored in the interchange record. This is of particular of interest in the case post-processing has been used (see clause 10). Then representations that are of type Post-processed shall code the ordinal number of the representation that they have been derived from, in the Cross Reference Field. The first representation of the interchange format has the code 0x01.

Table 21 — Cross Reference

Description	Value
Image Type is Basic, Token Frontal or Full Frontal	0x00
Image Type is Post Acquisition Processed	Ordinal number of the original representation

Example: There are four representations in the overall record. The second representation has been post-processed and resulted in the fourth representation. Then, the fourth representation shall have Cross Reference set to 2, all other Records shall have set Cross Reference to 0.

5.7.9 Image Colour Space

The (1 byte) Image Colour Space field indicates the colour space used in the encoded Image Information block according to the values in Table 22. The values of 0x80-0xFF are vendor specific. Application developers may obtain the values for these codes from the vendor.

Table 22 — Colour Space codes

Colour Space	Value
Unspecified	0x00
24 bit RGB	0x01
YUV422	0x02
8 bit greyscale	0x03
48 bit RGB	0x04
16 bit Greyscale	0x05
Other	0x06
Reserved by SC 37	0x07 – 0x7F
Vendor Specific	0x80– 0xFF

5.8 The Representation Data block

The Representation Data consists of the Image Data Block, the 3D Information Block and the 3D data Block.

5.9 The Image Data Block

5.9.1 Data structure

The (variable byte) Image Data block shall consist of two fields as shown in Table 23.

5.9.2 Image Data Length

This four byte field shall indicate the length of the image data in bytes.

5.9.3 Image data

This variable length field shall contain the image data encoded by the JPEG or JPEG2000 or PNG standards.

Table 23 —image data structure

Field	Size	Value	Notes
Length of Image data	4 bytes	$K \leq$ Length $\leq 2^{32} - 53$	Length of "Image data". K is minimum JPEG or JPEG 2000 or PNG header length.
Image data	Variable	See Table 18	Either JPEG or JPEG2000 or PNG

5.10 The 3D Information Block

The 3D Information block consists of the following fields and sub-blocks:

The Length of 3D Data Representation, the Coordinate System Type, the Texture Projection Matrix, Scale, Offset, the 3D Representation Type, the 3D Supplemental Data, a field reserved for future use, the 3D Capture Device Technology ID, the 3D Device Type, the 3D to 2D Image Temporal Synchronicity, the 3D to 2D-Texture Temporal Synchronicity, the 3D Acquisition Time, the 2D-Texture Acquisition Time, the Texture Map Type and finally the Texture Map Spectrum.

5.10.1 Length of 3D Data Representation

This (4 byte) field codes the length of the 3D Information and 3D Data block including the optional fields and blocks, if they are present.

5.10.2 Coordinate System Type

5.10.2.1 General

Originally, 3D data is acquired in a device dependent coordinate system. Based on the knowledge about several device parameters the 3D data can be transformed in metric world coordinates with two disadvantages:

- the regular structure of the device dependent data gets lost (e.g. leading to varying distances between data points)
- to obtain a regular structure in world coordinates, one has to interpolate. The original data is not preserved. This may be sufficient for many applications, but this standard is intended to be able to store the original data, too.

Thus the standard features several ways to store 3D data in different representations. All representations support a Cartesian coordinate system. The range data representation additionally supports a cylindrical coordinate system. Note that the coordinate system may be further restricted for different Face Image Types (see clauses 11 to 13, and Annex A.7).

The transformation to metric world coordinates is described by appropriate scaling factors and implicit rules (e.g. as used in the 3D anthropometric landmark type) defined in this standard (see clause 5.10.2.2 to 5.10.2.3),

The (1 byte) Coordinate System Type field specifies the coordinate system of the 3D data by using the following values.

Table 24 — The Coordinate System Type

Description	Value
Cartesian coordinate system	0x00
Cylindrical coordinate system	0x01
Reserved by SC 37	0x02 – 0xFF

The different coordinate systems are defined as follows:

5.10.2.2 Cartesian Coordinate System

In the Cartesian coordinate system the point of origin of the sensor data typically is used as the point of origin of the coordinate system.

The transformation from Cartesian coordinates to metric Cartesian coordinates is derived as follows:

$$X = x * ScaleX + OffsetX;$$

$$Y = y * ScaleY + OffsetY;$$

$$Z = z * ScaleZ + OffsetZ.$$

NOTE for certain Image Types the origin of the Cartesian coordinate system shall be the nose, i.e. the Prn landmark as defined in Table 16.

NOTE for certain Image Types the pose of the head is restricted. The frontal pose is defined by the Frankfurt Horizon FH (see Annex D) as the XZ plane and the vertical symmetry plane as the YZ plane with the Z axis oriented in the direction of the face sight.

A strong relation between anthropometric landmarks and the coordinate system is still established by

- the anatomical alignment requirements of the corresponding 2D image and
- the alignment between the 3D range data and the corresponding 2D image after applying the Texture Projection Matrix.

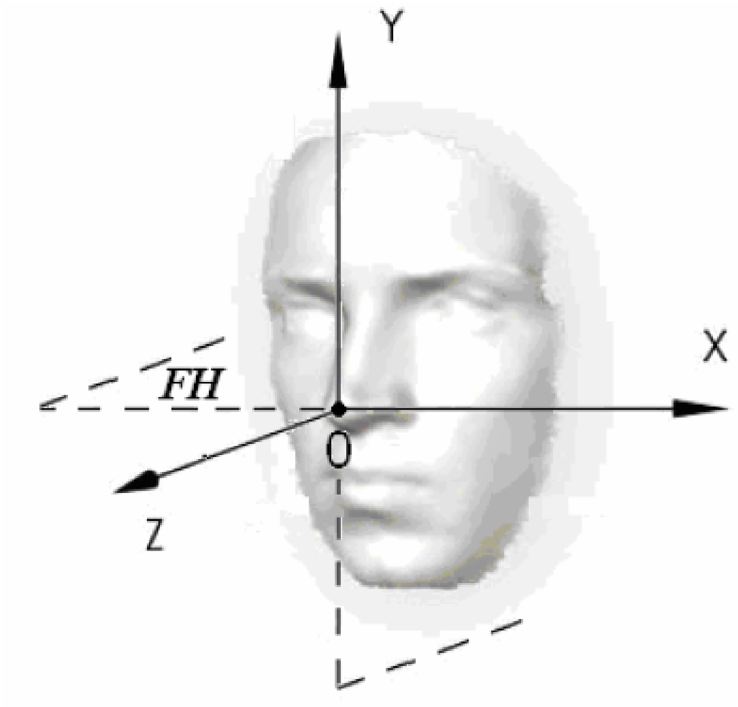


Figure 11 — A sample of a Cartesian coordinate system

5.10.2.3 The Cylindrical Coordinate System

A point in the Cylindrical Coordinate System is given by (α, h, r). The angle α and the h-axis are defined in a way that they form a clockwise coordinate system.

The transformation from cylindrical coordinates to metric Cartesian coordinates is derived as follows:

$$X = r * ScaleZ * \sin(\alpha * ScaleX) + OffsetX;$$

$$Y = h * ScaleY + OffsetY;$$

$$Z = r * ScaleZ * \cos(\alpha * ScaleX) + OffsetZ;$$

ScaleX, ScaleY, ScaleZ, OffsetX, OffsetY and OffsetZ are the necessary constants for the transformation. ScaleX has the physical unit of rad (degree radian). ScaleY, ScaleZ, OffsetX, OffsetY and OffsetZ are given in the physical unit mm (millimetre). Note, that large values of ScaleX, ScaleY or ScaleZ indicate a low spatial sampling rate in the respective dimension.

Typically, the point of origin of the sensor data is used as the point of origin of the cylindrical coordinate system.

NOTE for certain Image Types the origin of the cylindrical coordinate system shall be the nose, i.e. the Prn landmark as defined in Table 16.

A strong relation between anthropometric landmarks and the coordinate system is still established by

- the anatomical alignment requirements of the corresponding 2D image and
- the alignment between the 3D data and the corresponding 2D image after applying the Texture Projection Matrix.

The transformation from cylindrical coordinates to Cartesian coordinates is done by applying the transformation denoted in clause 5.10.2.3 and then inverting the transformation given in 5.10.2.2.

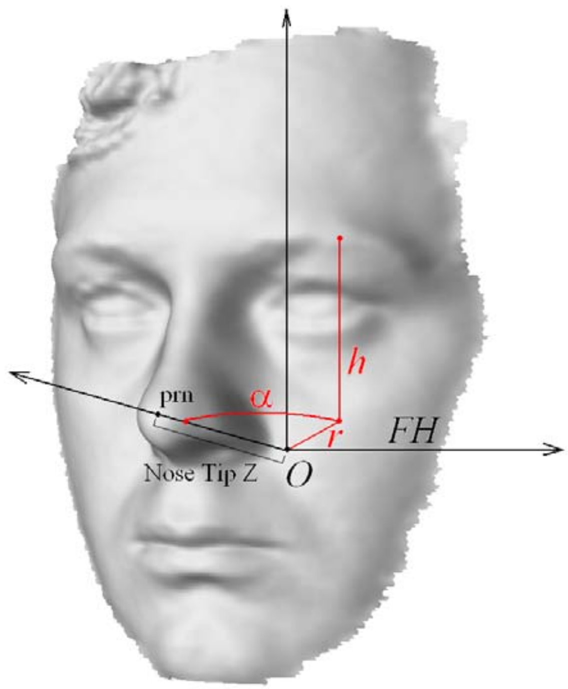


Figure 12 — A sample of a cylindrical coordinate system. FH is the Frankfurt Horizon as defined in Annex D.

5.10.3 Texture Projection Matrix

The Texture Projection Matrix P (3x4 float, 48 bytes) is required to map the 3D data onto the 2D texture image of the Image Data block. The matrix shall be stored row by row starting from the left top.

One can project a point in 3D space $[X,Y,Z]^T$ on the texture image of the Image Data block by multiplying the Texture Projection Matrix P with the so called homogeneous 3D coordinates of the 3D point [11].

$$[x,y,w]^T = P*[X,Y,Z,1]^T$$

Homogeneous 3D coordinates are a vector of four values $[X,Y,Z,1]^T$. Here X, Y, Z are the coordinates of a point in the metric cartesian coordinate system. The multiplication results in $[x,y,w]^T$, the so called homogeneous 2D coordinates with the auxiliary coordinate *w*. One obtains the resulting 2D image pixel coordinates of the texture image in the Image Data block by dividing the first two coordinates of the 2D homogeneous coordinates by the respective 3rd auxiliary coordinate *w*. Hence [x:w, y:w] are the resulting image pixel coordinates of the texture image related to the given 3D point $[X,Y,Z]^T$. Note, that the obtained coordinates are floating point values. In this standard there are no rules about how the necessary rounding or interpolation to the integer pixel coordinates has to be done.

In case the cylindrical coordinate system is used one shall transform to the metric Cartesian coordinate system to map the 3D data onto the texture. If there is overlapping, the texture is mapped to the first 3D point in the line of sight (closest to the observer).

The next two blocks store all necessary data to compute metric depth values from the 3D data.

5.10.4 ScaleX, ScaleY, ScaleZ, OffsetX, OffsetY, OffsetZ

As outlined in clause 5.10.2.2and 5.10.2.3 ScaleX, ScaleY, ScaleZ, OffsetX, OffsetY and OffsetZ are needed to transform digital coordinates to metric coordinates. This applies to all three 3D representations defined in this part of ISO/IEC 19794. The values are given in the physical unit mm (millimetre). In the case of Cartesian coordinates ScaleX also has the physical unit mm, in the case of a cylindrical coordinate system ScaleX has the physical unit of rad (degree radian). Each factor is represented by a mandatory four byte float value.

Large values of ScaleX, ScaleY or ScaleZ indicate a low spatial sampling rate in the respective dimension. Boundary values of ScaleX, ScaleY and ScaleZ may be strongly restricted for different Face Image Types (see clauses 11 to 13, and Annex A.7).

ScaleX and ScaleY in a range image represent sampling intervals while the ones in a 3D Point Map do quantization of the 3D space. Also, ScaleZ in either of these representations denotes quantization.

5.10.5 3D Representation Type

The (1 byte) 3D Representation Type shall be used to indicate the representation type that codes the 3D data.

Table 25 — 3D Representation Type

Description	Value
Range Image	0x00
3D Point Map	0x01
Vertex Data	0x02
Reserved for future definition by SC 37	0x03 - 0xFF

5.10.6 3D Supplemental Data

The (1 byte) 3D Supplemental Data mask is a bit mask of one byte and each bit of the mask position listed in Table 26 shall be set to 1 if the corresponding 3D information is present and set to 0 if absent. So, a bit mask of all zeros will indicate, that none of the options are present. The mask position starts from 0 at the lowest significant bit. The mask indicates if an Error Map/Vertex Error and/or a Texture Map is attached to the data.

Table 26 — 3D Supplemental Data

Description	Mask Position
Error Map or Vertex Error present	0
Texture Map present	1
Reserved for future definition by SC 37	2-7

5.10.7 3D Capture Device Technology ID

In analogy to the Capture Device Technology ID field in the 2D Image Information block, where the source of the 2D data can be coded, the (1 byte) 3D Capture Device Technology ID field should be used to indicate the type of the source that was used to acquire the 3D data. Additionally, the most significant bit (MSB) indicates if the scanning technology is active or passive for each source type.

Table 27 — The 3D Capture Device Technology ID

Description	Value (passive technology)	Value (active technology)
Unspecified	0x00	0x00
stereoscopic scanner	0x81	0x01
moving (monochromatic) laser line	Not available	0x02
structured light	Not available	0x03
colour coded light	Not available	0x04
ToF (Time of Flight)	Not available	0x05
Shape from Shading	0x86	0x06
Reserved by SC 37	0x87-0xFF	0x07 – 0x80

5.10.8 3D Device Type

The (2 byte) 3D Device Type field denotes the vendor specific capture device type ID. A value of all zeros will be acceptable and will indicate that the 3D Device Type is unspecified. Application developers may obtain the values for these codes from the vendor.

5.10.9 3D to 2D Image Temporal Synchronicity

The mandatory (2 byte) 3D to 2D Image Temporal Synchronicity shall be used to indicate the temporal relation between the 3D data and the 2D image of the Image Data block. It does **not** reference to the optional Texture Map of the 3D Data block. The value indicates the temporal difference between the start of the 2D and the start of the 3D acquisition process in milliseconds (ms). The field allows the coding of positive as well as negative differences. Here, a negative time difference denotes that the 3D acquisition started before the 2D acquisition. The time difference in milliseconds (ms) is coded in the two's complement system. So, a value of 0x8001 codes the maximum negative temporal difference of −32 767 ms and the value 0x7FFF corresponds to the maximum positive temporal difference of +32 767 ms. A value of 0x8000 is acceptable and indicates that the 3D Temporal Synchronicity is unspecified.

Table 28 — The 3D to 2D Temporal Synchronicity

Description	Value
Temporal difference between the start of the 2D and the 3D acquisition process in milliseconds (ms) in two's complement coding.	0x0000 – 0x7FFF 0x8001 – 0xFFFF
Unspecified	0x8000

5.10.10 3D to 2D Texture Temporal Synchronicity

The mandatory (2 byte) 3D to 2D Texture Temporal Synchronicity shall be used to indicate the temporal relation between the 3D data and the 2D textural data of the optional 2D Texture Map of the 3D Data block. The value indicates the temporal difference between the start of the Texture Map acquisition and the start of the 3D acquisition process in milliseconds (ms). NOTE It does **not** refer to the synchronicity between the acquisition of the 2D image in the Image Data block and the 3D data.

The field allows the coding of positive as well as negative differences. Here a negative time difference denotes that the 3D acquisition started before the 2D acquisition. The time difference in milliseconds (ms) is coded in the two's complement system. So, a value of 0x8001 codes the maximum negative temporal difference of −32 767 ms and the value 0x7FFF corresponds to the maximum positive temporal difference of +32 767 ms. A value of 0x8000 is acceptable and indicates that the 3D to 2D Texture Temporal Synchronicity is unspecified.

Table 29 — The 3D to 2D Texture Temporal Synchronicity

Description	Value
Temporal difference between the start of the optional 2D Texture Map and the 3D acquisition process in milliseconds (ms) in two's complement coding.	0x0000 – 0x7FFF 0x8001 – 0xFFFF
Unspecified	0x8000

5.10.11 3D Acquisition Time

Different 3D scanning techniques strongly vary in their acquisition time and this time may directly influence the quality of the data (if the subject moves during acquisition). Therefore, the (2 byte) 3D Acquisition Time field is used to code the time span between the start of the 3D acquisition process and the end of the 3D acquisition process in ms (milliseconds). A value of 0xFFFF is acceptable and indicates that the field is not specified.

Table 30 — The 3D Acquisition Time

Description	Value
Duration of the 3D acquisition process in milliseconds (ms)	0x0000 – 0xFFFE
Unspecified	0xFFFF

5.10.12 2D Texture Acquisition Time

The optional 2D Texture Map of the 3D record may or may not be simultaneously acquired with the 3D data. Therefore, the (2 byte) 2D Texture Acquisition Time field is used to code the time span between the start of the 2D acquisition process and the end of the 2D acquisition process of the optional Texture Map in ms (milliseconds). A value of 0xFFFF is acceptable and indicates that the field is not specified. NOTE this is **not** the time needed to acquire the 2D image of the Image Data block.

Table 31 — The 2D Texture Acquisition Time

Description	Value
Duration of the 2D acquisition process in milliseconds (ms)	0x0000 – 0xFFFE
Unspecified	0xFFFF

5.10.13 Texture Map Type

The (1 byte) Texture Map Type field denotes the encoding type of the Texture Map block. If the 3D Supplemental Data field specifies that there is a Texture Map in the record, either JPEG (ISO/IEC 10918-1 and ITU-T rec. T.81) or JPEG 2000 (ISO/IEC 15444-1) or PNG (ISO/IEC 15948:2004) shall be specified. For JPEG, the data shall be formatted in accordance with the JPEG File Interchange Format, Version 1.02 (JFIF).

If the 3D Supplemental Data field specifies that there is no Texture Map in the record the Texture Map Type shall be “Unspecified”.

Table 32 — The Texture Map Type codes

Description	Value
Unspecified	0x00
JPEG	0x01
JPEG2000	0x02
PNG	0x03
Reserved by SC 37	0x04-0xFF

5.10.14 Texture Map Spectrum

The (1 byte) Texture Map Spectrum field denotes the kind of spectrum that has been used for acquiring the Texture Map specified in clause 5.11.9. Whereas the 2D face image always uses the spectrum of the visible light, this can be different for the acquisition of the Texture Map. If the 3D Supplemental Data field specifies that there is a Texture Map in the record, the Texture Map Spectrum field shall not be unspecified.

If the 3D Supplemental Data field specifies that there is no Texture Map in the record, the Texture Map Spectrum field shall be unspecified.

Table 33 — The Texture Map Spectrum codes

Description	Value
Unspecified	0x00
Visible (380nm- 780nm)	0x01
Very-near infrared (photographic) (780nm-1000nm)	0x02
Short wave infrared (1000nm-1400nm)	0x03
Other	0x04
Reserved by SC 37	0x05-0xFF

5.11 The 3D Data Block

5.11.1 Data Structure

The 3D Data block contains the representation of the 3D data. There are three alternatives to store 3D data: a Range Image, 3D Point Map, or using the Vertex representation. Optionally, additional information can be stored in the Error Map and Vertex Error, respectively, and in the Texture Map.

Range Images are well suited for the coding of the depth values of a specific view of an object projected on a plane or a cylinder. By definition, only one depth value per pixel can be coded, which restricts the complexity of the coded surface. Nevertheless, for facial images, esp. frontal facial images, this typically is a very good approximation. Range Images may be less suited for coding of depth information showing strong poses. Furthermore, depth information in Range Images is typically more processed (smoothed, re-sampled, interpolated, etc.) than data in the 3D Point Map.

The 3D Point Map is most suited for exchange and storage of raw, unprocessed 3D sensor data. Storing raw data may result in larger sizes of the representation.

Vertex data codes 3D points based on a non-regular sampling interval, typically resulting in a sparse coding. Due to the variable sampling of the vertex points the Vertex data representation on the one hand side can result in very compact representations or in a very exact representation when using many vertices.

The 3D Representation Type field (see clause 5.10.5) is used to define the format of the 3D data representation that has been used in the actual record.

5.11.2 Range Image Bit Depth

The (1 byte) Range Image Bit Depth field denotes the bit depth of the Range Image. This field is given for the sake of easier record parsing, as the bit depth can also be derived from the PNG record header.

Table 34 — The Range Image Bit Depth codes

Description	Value
8 bit	0x00
16 bit	0x01
Reserved by SC 37	0x02-0xFF

5.11.3 Range Image

The Range Image is a representation of the range data in a two dimensional form. The Range Image shall be stored in the PNG format (ISO/IEC 15948:2004). PNG provides lossless compression for both 8 and 16-bit grey scale representation data. The bit rate of the PNG code is written in the PNG header, but shall also be given in the Range Image Bit Depth field (see clause 5.11.2). Hence whether the 8 or 16 bit depth coding is used shall be defined from the PNG record header.

The length of the map is variable, as it depends on the lossless compression algorithm. The uncompressed data has the dimension Range Image Height x Range Image Width. These dimensions are encoded in the PNG header.

Pixel of value 0xFF for 8 bit PNG coding and 0xFFFF for 16 bit PNG coding shall indicate non-valid range data.

5.11.4 3D Point Map Width and Height

These two fields define the width and height of the 3D Point Map where the 3D data is stored. Both fields are 2 byte values ranging from 0 to 65 535.

5.11.5 3D Point Map

The 3D Point Map allows storing of raw 3D scanner data. The organization of this block is as follows. It consists of a three channel lossless compressed image in the PNG format with 16 bits per channel. The first channel represents the X, the second the Y, and the third the Z values. A pixel value of (X,Y,Z) = (0xFFFF, 0xFFFF , 0xFFFF) shall be used to indicate a non-valid 3D point.

The coordinates are given in an arbitrary cartesian coordinate system. Connectivity is not explicitly encoded. For valid points neighbouring pixel positions represent neighbouring positions on the face surface.

5.11.6 Vertex Data

The variable length Vertex Data block contains the Vertex Coordinates block, the optional Vertex Normals block, the optional Vertex Errors block, and the optional Vertex Textures block. Each of these blocks contains a list of vertex descriptions. The number of the vertex descriptions is given by the (2 byte) Vertex Count field.

The location of each vertex is represented by its X coordinate, Y coordinate, and Z coordinate as specified in the (2 byte) Vertex X Coordinate, Vertex Y Coordinate and Vertex Z Coordinate fields, respectively. The values code the location with a fixed precision as specified in clause 11.3.2.

If the Normal Flag is equal to 0x01 the corresponding normal vector to each vertex shall be specified in the (2 byte) Normal X, Normal Y and Normal Z Coordinate fields, respectively.

The optional (1 byte) Vertex Error field codes additional information on the vertex as described in Table 36 in clause 5.11.8 of this standard. If the existence of an Error Map is specified in the 3D Supplemental Data field, the Vertex Error field shall be present for each vertex.

The optional Vertex Texture X and Vertex Texture Y fields represent the corresponding x and y pixel position in the Texture Map with (0,0) denoting the upper left corner. If the existence of a Texture Map is specified in the 3D Supplemental Data field, Vertex Texture X and Vertex Texture Y shall be present for each vertex.

The number of triangles is specified in the (4 byte) Triangle Face Count field.

The 3D Vertex Data representation optionally allows for the specification of additional normals to the vertexes. This can be indicated by the (1 byte) Normal Flag field.

Table 35 — The Normal Flag codes

Description	Value
Normal information not used in Vertex Data	0x00
Normal information used in Vertex Data	0x01
Reserved by SC 37	0x02-0xFF

5.11.7 Triangle Data

The variable length Vertex Triangle Data contains a list of triangle descriptions. The number of the triangle descriptions is given by the Triangle Face Count field (see clause 5.11.6). Each triangle is specified by the three (2 byte) indices of the vertices in the vertex data list forming the triangle. The order of the vertex indices shall be counter clock wise to indicate the external face of the triangle.

5.11.8 Error Map

The optional Error Map can be used to further give information on how the 3D data has been processed before it was stored in the 3D representation. The Error Map shall be coded in the PNG format using an 8bit per pixel greyscale image. The length of the map is variable, as it depends on the lossless compression algorithm. The uncompressed data has the dimension Range Image Height x Range Image Width in the case it is associated with a Range Image or 3D Point Map Width x 3D Point Map Height in case it is associated with a 3D Point Map.

Pixel values t in the range of 0 to 199 are reserved for future use by SC 37. A value of $t = 200$ code that the depth value is considered to be correct. Values of $t = 201$ and above code a specific potential or corrected defect of the 3D data or the corresponding Texture Image.

See clause 5.11.6 of how the pixel values are used in the 3D Vertex representation.

Table 36 — The Error Map

Description	Value
Reserved for future use by SC 37	0…199
Depth value is considered correct	200
Depth value is interpolated, interpolation type isn't specified	201
Depth value is interpolated, linear interpolation has been used	202
Depth value is interpolated, bicubic interpolation has been used	203
Value of optional texture image potentially wrong (texture noisy, overexposure, etc.)	204
Value of optional texture image has been corrected by post processing (image processing)	205
Reserved for future use by SC 37	206…255

5.11.9 Texture Map

The optional Texture Map should only be used to store textural face data that is acquired by a scanning device during the 3D acquisition process, and therefore may have geometry other than the standard 2D image stored in the Image Data block of the same record. It is not a substitute for the mandatory 2D image of the Image Data block. The Texture Map has the format specified in the Texture Map Type field. It can be coded in 8 bit or 16 bit greyscale or 24 bit colour image. The length of the map is variable as it depends on the compression algorithm. The uncompressed data has the dimension Range Image Height x Range Image Width in the case it is associated with a Range Image or 3D Point Map Width x 3D Point Map Height in case it is associated with a 3D Point Map, and variable dimensions when associated with a 3D Vertex representation.

6 The Basic Face Image Type

6.1 Inheritance requirements for the Basic Face Image Type

The Basic Face Image Type is the base class of all Face Image Types. All Face Image Types obey normative requirements of this clause (6). The inheritance map for Image Types is shown in Figure 10.

6.2 Image data encoding requirements for the Basic Face Image Type

One of three possible encodings is to be used for all image types (Basic)

1) The JPEG Sequential baseline (ISO/IEC 10918-1) mode of operation and encoded in the JFIF file format (the JPEG file format)

2) The JPEG-2000 Part-1 Code Stream Format (ISO/IEC 15444-1) and encoded in the JP2 file format (the JPEG2000 file format)

3) The PNG (ISO/IEC 15948:2003) standard. PNG shall not be used in its interlaced mode and not for images that have been JPEG compressed before.

6.3 Image data compression requirements for the Basic Face Image Type

Both encoding methods allow for compression of image data. There are no normative requirements on compression for the Basic Face Image Type. Compression is discussed further in informative annex A.1.

6.4 Format requirements for the Basic Face Image Type

6.4.1 Facial Header

The Format Identifier, Version Number, Length of Record, and Number of Representations fields shall be specified.

6.4.2 Facial Information

The Representation Length and Number of Landmark Points fields shall be specified.

6.4.3 Image Information

The Face Image Type field shall be specified with value 0x00.

The Image Data Type, Width, and Height fields shall be specified.

7 The Frontal Face Image Type

7.1 Inheritance requirements for the Frontal Face Image Type

The Frontal Face Image Type is a subclass of the Basic Face Image Type and therefore obeys all normative requirements of clause 6 “The Basic Face Image Type”. Note that the Frontal Face Image Type is not a valid Face Image Type but helps to describe common specifications of the Full Frontal Image Type and the Token Image Type. Therefore, all Frontal Type Images have to be either Full Frontal or Token Frontal. (Refer clause 5.7.2)

7.2 Scene requirements for the Frontal Image Type

7.2.1 Purpose

This clause specifies scene constraints for the capture of Frontal images, of either Image Type Full Frontal or Token. This clause should be read in conjunction with Informative Annex A.2 “Best Practices for Frontal Images”.

7.2.2 Pose

Pose is known to strongly affect performance of automated face recognition systems. Thus, the full-face frontal pose shall be used. Rotation of the head shall be less than +/- 5 degrees from frontal in pitch and yaw (ref. 5.5.8). Pose variations that lead to an in-plane rotation of the head can be more easily compensated by automated face recognition systems. Therefore, the rotation of the head shall be less than +/- 8 degrees from frontal in roll (ref. 5.5.8). Figure 13 shows an example of +/-8 degree rotation in roll.

Figure 13 — Sample images with +8 degrees (left) and -8 degrees (right) rotation in roll

NOTE It is noted that zero degree pitch angle is not defined in this standard explicitly.

This constraint refers to the pose of the subject associated with the Face Image format data for all applications that call for this format to be used.

7.2.3 Expression

Expression is known to strongly affect the performance of automated face recognition systems. It is recommended to classify the expression as one of the following.

a) Neutral (non-smiling) with both eyes open normally (i.e. not wide-open), and mouth closed.

b) A smile where the inside of the mouth and/or teeth is not exposed (closed jaw).

c) A smile where the inside of the mouth and/or teeth is exposed.

d) Raised eyebrows

e) Eyes looking away from the camera

f) Squinting

g) Frowning

See informative annex A.2.2 for best practices on this topic based upon this classification scheme.

7.2.4 Assistance in positioning the face

In no cases will any other face be captured in the Frontal image. See informative annex A.2 for best practices on this topic.

7.2.5 Shoulders

Shoulders shall be “square on” to the camera. “Portrait style” photographs where the subject is looking over one shoulder are not acceptable.

7.2.6 Backgrounds

Specification of background is not normative for the creation of Frontal images. See informative annex A.2 for best practices on this topic.

7.2.7 Subject and scene lighting

Lighting shall be equally distributed on the face. There shall be no significant direction of the light from the point of view of the photographer, as further described in clauses A.2.7 and A.2.8. The ratio between the median intensity on a square region centred around Landmark Points 5.3 and 5.4 with side length 20% of the inter-eye distance shall be between 0,5 and 2,0.

7.2.8 Hot spots and specular reflections

Care shall be taken to avoid Hot spots (i.e., bright regions that results from light shining directly on the face) shall be absent. These artefacts typically occur when a high intensity focused light source is used for illumination. Diffused lighting, multiple balanced sources or other lighting methods shall be used.

NOTE Implementers might refer to Annexes A and B which advance best practice.

NOTE A single bare “point” light source like a camera mounted flash is not acceptable for imaging. Instead, the illumination should be accomplished using other methods that meet requirements specified in this clause.

7.2.9 Eye glasses

Glasses should be clear glass and transparent so that the eye pupils and irises shall be visible. This requirement is intended to exclude dark or otherwise opaque glasses. Tinted glasses or sunglasses shall not be worn. An exception applies when the subject asserts a medical reason to retain the glasses; in these cases the dark glasses indicator in the header structure shall be set.

The frames of glasses shall not obscure the eyes. The frames shall not be thicker than 5% of the distance between points 12.1 and 12.2 (midpoints of left and right eye) in Figure 8.

NOTE Refer to Annex A.2.5 for additional information on this topic.

7.2.10 Head coverings

In cases head coverings are present the related flag in the property mask shall be set.

7.2.11 Visibility of pupils and irises

In cases pupils or irises are not visible the related flag in the property mask shall be set.

7.2.12 Lighting artefacts

There shall be no lighting artefacts or flash reflections on glasses. Lighting artefacts covering any region of the eyes shall not be present. This applies to any region in polygon between landmark points 3.8, 3.2, 3.12 and 3.4 for the right eye and between landmark points 3.11, 3.1, 3.7 and 3.3 for the left eye in Figure 7.

7.2.13 Eye patches

Eye patches shall not be worn. An exception applies when the subject asserts a need to retain the patch (e.g. a medical reason) is claimed; in these cases the left or right patch indicators in the header structure shall be set.

7.3 Photographic Requirements for the Frontal Image Type

7.3.1 Purpose

This clause specifies photographic constraints for the capture of Frontal face images, of either type Full Frontal or Token. Rather than impose a particular hardware and lighting capture system, this clause specifies the type of output from these systems that is allowed. This clause applies to film as well as digital photography, and it should be read in conjunction with Informative Annex A.2 “Best Practices for Frontal Images”.

7.3.2 Contrast and saturation

For each patch of skin on the person’s face, the gradations in textures shall be clearly visible i.e. being of reasonable contrast. In this sense, there will be no saturation (over or under exposure) on the face.

7.3.3 Focus and depth of field

The subject’s captured image shall always be in focus from nose to ears and chin to crown. Although this may result in the background behind the subject being out of focus, this is not a problem.

NOTE In a typical photographic situation, for optimum quality of the captured face, the f-stop of the lens should be set at two (or more) f-stops below the maximum aperture opening when possible to obtain enough depth of field.

All images shall have sufficient depth of focus to maintain greater than two millimetre spatial sampling rate on the subject’s facial features at time of capture.

Further information on depth of field is included in Annex B.2.1.1 .

7.3.4 Unnatural colour

Unnaturally coloured lighting, yellow, red, etc. is not allowed. Care shall be taken to correct the "white balance" of image capture devices. The lighting shall produce a face image with natural looking flesh tones when viewed in typical examination environments. Images showing the “red-eye effect”, i.e. the common appearance of red eyes on photographs taken with a photographic flash when the flash is too close to the lens, are not acceptable. The iris and the iris colour shall be visible.

NOTE see Annex B for best practices related to this topic.

7.3.5 Colour or greyscale enhancement

A process that overexposes or under-develops a colour or greyscale image for purposes of beauty enhancement or artistic pleasure is not allowed. The full spectrum shall be represented on the face image where appropriate. Teeth and whites of eyes shall be clearly light or white (when appropriate) and dark hair or features (when appropriate) shall be clearly dark.

7.3.6 Radial distortion of the camera lens

The fish eye (ref. 4.11) that is associated with unusually large noses in the image is not allowed.

NOTE While some distortion is almost always present during portrait photography, that distortion should not be noticeable by human examination. See informative annex A.2 for further discussion.

7.4 Digital requirements for the Frontal Image Type

This clause discusses normative aspects of the digital properties of Frontal Images including Full Frontal and Token.

7.4.1 Geometry

7.4.1.1 Pixel aspect ratio

Digital cameras and scanners used to capture facial images shall produce images with a pixel aspect ratio of 1:1. That is, the number of pixels per inch in the vertical dimension shall equal the number of pixels per inch in the horizontal direction.

7.4.1.2 Origin at upper left

The origin of coordinates shall be at the upper left given by coordinate (0,0) with positive entries from left to right (first dimension) and top to bottom (second dimension).

7.4.2 Colour profile

7.4.2.1 Colour space

Frontal images shall be represented as one of the following

a) The 24-bit RGB colour space where for every pixel, eight (8) bits will be used to represent each of the Red, Green, and Blue components.

b) An 8-bit monochrome colour space where for every pixel, (8) bits will be used to represent the luminance component.

c) The YUV422 colour space where twice as many bits are dedicated to luminance as to each of the two colour components. YUV422 images typically contain two 8-bit Y samples along with one 8-bit sample of each of U and V in every four bytes.

NOTE To achieve device-independence, the RGB values from the camera or scanner should be converted to values in a defined standard RGB space, such as sRGB [19], using the device's colour profile and colour management processing. Information regarding device profiling and colour management can be downloaded from the International Color Consortium URL: www.color.org.

7.4.3 Video interlacing

Interlaced video frames are not allowed for the Frontal Image Type. All interlacing must be absent (not simply removed, but absent).

7.4.4 Use of near infra-red cameras

Dedicated near infra-red cameras shall not be used for acquisition of image of the Frontal Image Type.

7.5 Format requirements for the Frontal Image Type

7.5.1 Inheritance requirements

The format requirements for the Basic Face Image Type shall be specified, as given in clause 6.4. In addition the following requirements shall be specified.

7.5.2 Image Information

Frontal Images are either Full Frontal or Token Frontal images and the Face Image Type field shall be set accordingly (ref. 8.5.2, 9.3.2).

8 The Full Frontal Image Type

8.1 Inheritance requirements for the Full Frontal Face Image Type

The Full Frontal Face Image Type is a subclass of the Frontal Image Type and therefore obeys all normative requirements of clause 6 “The Basic Face Image Type” and clause 7 “The Frontal Face Image Type”.

8.2 Scene requirements for the Full Frontal Face Image Type

The Full Frontal Face Image Type is a subclass of the Frontal Image Type and therefore obeys all normative requirements of clause 6 “The Basic Face Image Type” and clause 7 “The Frontal Face Image Type”.

8.3 Photographic requirements for the Full Frontal Face Image Type

8.3.1 Introduction

This clause describes the *minimum* relative dimensions of the full image with respect to the face. The requirements of this clause can be met by images taken in both portrait and landscape mode, and Figure 14 shows a portrait image and head outline to display dimensions A, B, BB, CC, and DD which are referenced in sub-clauses below. In addition to the requirements below, the face from chin to crown as defined in 8.3.5 and with the full width as defined in 8.3.4, shall be visible in the image.

Informative annex A.3.2 discusses additional constraints on image and head dimensions for sizes appropriate specifically to travel documents.

Note that for digital images the normative requirements related to the minimum inter-eye distance as defined in clause 8.4.1 impose further requirements on the minimum head size. See Annex A.3.1.1 for more information regarding the connection between photo spatial sampling rate and the photographic requirements of this clause.

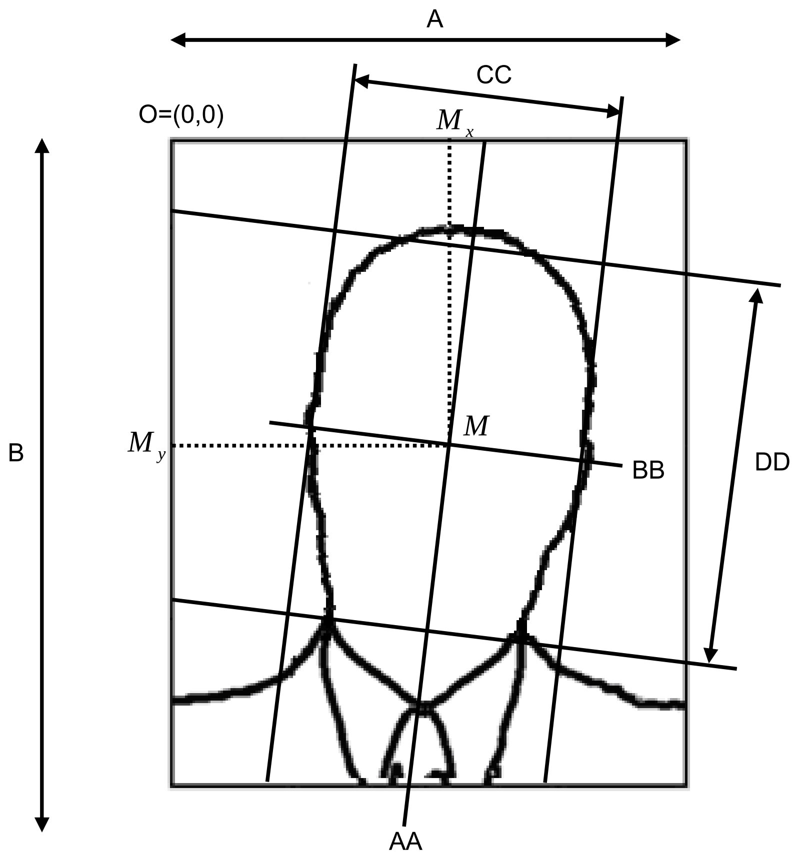


Figure 14 — Geometric characteristics of the Full Frontal Face image

8.3.2 Horizontally centred face

The approximate horizontal midpoints of the mouth and of the bridge of the nose define the imaginary line AA (usually the symmetry axis of the face) . Furthermore, the imaginary line BB is defined as the line through the centres of the left and the right eye. The intersection of AA and BB defines the point M as the centre of the face. The X coordinate M_x of M shall be between 45% and 55% of the image width.

8.3.3 Vertical position of the face

The Y coordinate M_y of M shall be between 30% and 50% of the image height. A single exception is allowed for children under the age of 11 years, in which case the higher limit shall be modified to 60% (i.e. the centre point of the head is allowed to be lower in the image for children under the age of 11). Note that the origin O of the coordinate system is in the upper left corner of the image.

8.3.4 Width of head

The width of a head is defined as the distance between the two imaginary lines parallel to the line AA; each imaginary line is drawn between the upper and lower lobes of each ear and shall be positioned where the external ear connects the head. The head width is shown as length CC in Figure 14[2].

To ensure that the entire face is visible in the image the head width CC shall be between 50% and 75% of the image width (A).

---

[2] This figure is a derivative of AAMVA document DL/ID-2000.

8.3.5 Length of head

The length of a head is defined as the distance between the base of the chin and the crown measured on the imaginary line AA. This is shown as length DD in Figure 14. The crown is defined as the top of the head ignoring any hair.

In order to assure that the entire face is visible in the image, the minimum image height shall be specified by requiring that the crown-to-chin portion (DD) of the Full Frontal image pose shall be between 60% and 90% of the vertical length of the image (B). A single exception is allowed for children under the age of 11 years, in which case the lower limit shall be modified to 50%.

8.3.6 Summary of photographic requirements

Table 37 below summarizes the photographic requirements for Full Frontal images specified in clauses 8.3.1–8.3.5.

Table 37 — Summary of photographic requirements for Full Frontal Images

Clause	Definition	Requirements
8.3.1	General requirement	Head entirely visible in the image
8.3.2	Horizontal Position of Face	0.45 A ≤ M_x ≤ 0.55 A
8.3.3	Vertical Position of Face	0.3 B ≤ M_y ≤ 0.5 B
8.3.3	Vertical Position of Face (Children under the age of 11)	0.3 B ≤ M_y ≤ 0.6 B
8.3.4	Width of Head	0.5 A ≤ CC ≤ 0.75 A
8.3.5	Length of Head	0.6 B ≤ DD ≤ 0.9 B
8.3.6	Length of Head (Children under the age of 11)	0.5 B ≤ DD ≤ 0.9 B

Figure 15 shows a typical example of a passport image. The outer rectangle visualizes the maximum dimensions of the head based on the requirements in clause 8.3.4 and 8.3.5. Furthermore, the inner rectangle shows the minimum width and height dimensions of the head based on the image dimensions.

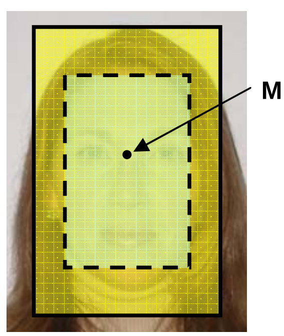


Figure 15 — A sample image with the respective minimal and maximal head dimensions based on the image width and height

8.4 Digital requirements for the Full Frontal Face Image Type

8.4.1 Spatial Sampling Rate

For an image for optimal human examination and permanent storage, the spatial sampling rate of the full images shall be at least 180 pixels of spatial sampling for the width of the head, or roughly 90 pixels from eye centre to eye centre. See informative annex clause A.3.1.1 for best practices on this topic.

8.4.2 Post-acquisition processing

No other post processing than in-plane rotation and/or cropping and/or down sampling and/or multiple compression shall be applied to derive a Full Frontal Face image from a captured image**.**

8.5 Format requirements for the Full Frontal Image Type

8.5.1 Inheritance requirements

The format requirements for the Basic Face Image Type shall be specified, as given in clause 6.4. In addition the following requirements (clause 8.5.2) shall be specified.

8.5.2 Image Information

The Face Image Type field shall be specified with value 1.

9 The Token Face Image Type

9.1 Inheritance requirements for Token Face Image Type

The Token Face Image Type is a subclass of the Frontal Image Type and therefore obeys all normative requirements of clause 6 “The Basic Face Image Type” and clause 7 “The Frontal Face Image Type”.

9.2 Digital requirements for the Token Face Image Type

9.2.1 Introduction

The Token Face image is used to store the extracted face information from any other image source. The Token Face image inherits properties from the Frontal Face image format.

It can be generated at any resolution using only the pixel positions of the centre of the eyes relative to the upper left corner of the full image. The purpose of the Token Face image is to standardize the position of the eyes in an image and define the minimal amount of image area around the eyes. Using a Token Face image representation may help to reduce the amount of data stored for facial images while retaining the information needed for automated face recognition applications.

9.2.2 Eye positions

To create a Token Face image, the eye socket centres, or simply eye positions, defined as Landmark Points 12.1 and 12.2, shall be determined. For the determination of eye positions, it is possible:

1 - to use computer inspection

2 - to use human visual inspection

3 - to use computer and human visual inspection.

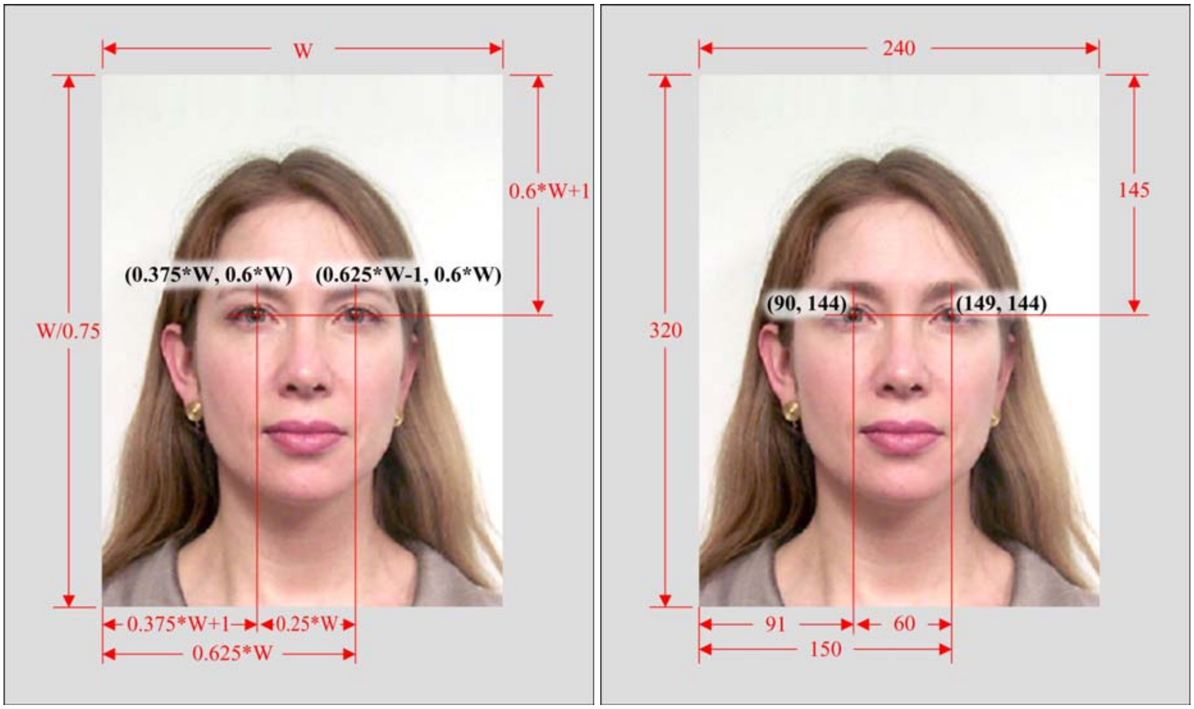
9.2.3 Geometric characteristics

A Token image is a colour or greyscale image with image dimensions and eye position coordinates given by Table 38. Note that clause 5.2.3 specifies conversion of values to integer.

Table 38 — The geometric characteristics of the Token Image Type

Feature or Parameter	Value
Image Width	W
Image Height	W / 0,75
Y coordinate of Eyes	0,6 * W
X coordinate of First (right) Eye	0,375 * W
X coordinate of Second (left) Eye	(0,625 * W) - 1
Width from eye to eye (inclusive)	0,25 * W

An example is shown in Figure 16.



a) b)

Figure 16 – The geometric requirements of the Token Face Image Type (a) and a sample image of the Token Face Image format showing the minimum required image width W=240 pixels (b)

9.2.4 Minimum width of the Token Image Type

The minimum required image width is 240 pixels. This corresponds to an image height of 320 pixels, a Y coordinate of eyes of 144, X coordinate of the first eye of 90 and X coordinate of second eye of 149. The distance from eye to eye (inclusive) in this case is therefore 60 pixels. This example is shown in Figure 16.

Coordinates are relative to the top left corner of the image (0,0) and all measurements are in units of pixels.

9.2.5 Padding

The normative practice shall be to fill any undefined set of pixels with any colour. See informative annex A.4.3 for best practices on this matter.

9.2.6 Post-acquisition processing

No other post processing than in-plane rotation and/or cropping and/or down sampling and/or multiple compression shall be applied to derive a Token Frontal Face image from a captured image.

9.3 Format requirements for the Token Face Image Type

9.3.1 Inheritance requirements

The format requirements for the Basic Face Image Type shall be specified, as given in clause 6.4. In addition the following requirements shall be specified.

9.3.2 Image Information

The Face Image Type field in the Image Information structure shall be specified with value 0x02.

10 The Post-processed Frontal Face Image Type

10.1 Introduction

Applying digital post-processing to a captured image can be changed in a way, that it is more suitable for automatic face recognition. The Post-processed Frontal Face Image Type is thought to exchange these kinds of facial images.

10.2 Inheritance requirements for the Post-processed Frontal Face Image Type

The Post-processed Frontal Face Image Type is a subclass of the Frontal Image Type and therefore obeys all normative requirements of clause 6 “The Basic Face Image Type” and clause 7 “The Frontal Face Image Type”.

10.3 Format requirements for the Post-processed Frontal Face Image Type

10.3.1 Inheritance requirements

The format requirements for the Basic Face Image Type shall be specified, as given in clause 6.4. In addition the following requirements shall be specified.

10.3.2 Block referencing

As biometric matchers may be very sensitive to artefacts caused by post-processing the parent non-post-processed image (e.g. in form of a Full Frontal, Token Frontal or Basic Image) should be stored in addition to the processed image in another representation for interoperability reasons.

Also if the receiver of an interchange record containing post-processed data does have better technology than the encoder of this record they can still rely on the original data.

To encode the relationship between the different representations, the Cross Reference field shall be used, i.e. the Cross Reference field shall not be zero. Multiple cross referencing of one original representation is possible.

10.3.3 Image Information

The Face Image Type field in the Image Information structure shall be specified with value 0x03.

10.3.4 Post-acquisition processing

The post acquisition processing bit-field shall be specified, i.e. its value shall be greater than 0.

11 The Basic 3D Image Type

11.1 Inheritance Requirements for the Basic 3D Image Type

The Basic 3D Image Type is the base class of all 3D Face Image Types. All 3D Face Image Types obey normative requirements of this clause (11).

Furthermore, the Basic 3D Image Type inherits all the requirements of the Basic Face Image type.

All mandatory (non-optional) fields of the 3D Information block shall be defined. Note that some of the mandatory fields may still remain unspecified if the appropriate value is set. Please see the specification of the individual fields for details.

11.2 The Basic 3D Image Type using the 3D Point Map representation

11.2.1 Coordinate System Type

The Coordinate System Type for the Basic 3D Image Type using the 3D Point Map representation shall be 0x00, i.e. a Cartesian coordinate system shall be used.

11.2.2 ScaleX, ScaleY and ScaleZ

Basic 3D Images using the 3D Point Map representation shall use a fixed scaling and offset values. The following values shall be used:

ScaleX = ScaleY = ScaleZ = 0,02 mm

OffsetX = OffsetY = OffsetZ = -655,34mm

11.3 The Basic 3D Image Type using the 3D Vertex representation

11.3.1 Coordinate System Type

The Coordinate System Type for the Basic 3D Image Type using the 3D Vertex representation shall be 0x00, i.e. a Cartesian coordinate system shall be used.

11.3.2 ScaleX, ScaleY and ScaleZ

Basic 3D Images using the 3D Vertex representation shall use a fixed scaling and offset values. The following values shall be used:

ScaleX = ScaleY = ScaleZ = 0,02 mm

OffsetX = OffsetY = OffsetZ = -655,34mm

12 The Full Frontal 3D Image Type

The Full Frontal with 3D Image Type shall fulfil the following requirements.

12.1 Inheritance requirements

The Full Frontal with 3D Image Type inherits the requirements of the Basic 3D Image Type as specified in clause 11.1. Furthermore, it inherits all requirements of the Full Frontal Image Type.

12.2 Coordinate System Type

The Coordinate System Type for the Full Frontal 3D Image Type shall be 0x00, i.e. a Cartesian coordinate system shall be used. Furthermore, the origin of the coordinate system shall be the nose, i.e. the Prn landmark as defined in Table 16.

12.3 Pose of the 3D representation

The rotation of the head in the 3D representation shall be less than +/- 5 degrees from frontal in pitch and yaw (see clauses 5.5.8.2 and 5.5.8.1). Pose variations that lead to an in-plane rotation of the head can be more easily compensated by automated face recognition systems. Therefore, the rotation of the head shall be less than +/- 8 degrees from frontal in roll (ref. clause 5.5.8.3). This constraint refers to the pose of the subject associated with the Face Image Format data for all applications that call for this format to be used.

12.4 Calibration Texture Projection Accuracy

The calibration accuracy of the acquisition device shall be so high, that the mean shift between the texture of the 2D Full Frontal Image and the 3D data after projection with the texture projection matrix is less than 1mm. Note: This may not represent accuracy obtained during normal usage due to subject/device movement. The 3D to 2D Image Temporal Synchronicity (see clause 5.10.9) will be indicative of the observed effect.

12.5 Requirements on Full Frontal 3D Image Types using the Range Image Representation

12.5.1 ScaleX, ScaleY and ScaleZ

The resolution of the stored depth data strongly depends on ScaleZ. To preserve quality a maximum value of $ScaleZ_{\max} = 1mm$ shall be defined for the Full Frontal 3D Image Type.

For the same reason a maximum value of $ScaleX_{\max} = ScaleY_{\max} = 1mm$ shall be defined for the Full Frontal 3D Image Type in a Cartesian coordinate system.

Note, that in any case, ScaleX and ScaleY denote a sampling rate, not the physical measurement rate of the sensor.

12.5.2 Face Coverage

The 3D data shall cover the minimal rectangular dimensions [-1,75 *w*, 1,75 *w*] x [-1,75 *w*, 2,55 *w*] ("outer region") in the Cartesian Coordinate System with the Landmark Point Prn as origin, where *w* is the distance between the feature points 12.1 and 12.2 (centre of the eyes) as defined in clause 5.6.4. Figure 17 shows a sample 2D image with the highlighted outer region.

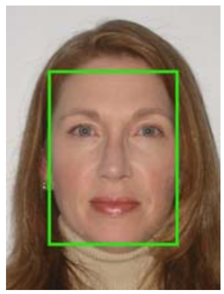


Figure 17 – Sample 2D image of the minimal Face Coverage (outer region).

12.5.3 Non-valid points in 3D data Image

At maximum 50% of the pixel of the Range Image in the region defined in clause 12.5.2 shall have a zero value, indicating a non-valid depth value. Furthermore, in the "inner region" defined as [-1,5 *w*, 1,5 *w*] x [-1,8 *w*, 1,8 *w*] in the Cartesian coordinate system with the Landmark Point Prn as origin at maximum 20% of the pixel shall have a zero value, indicating a non-valid depth value. Here, *w* is the distance between the feature points 12.1 and 12.2 (centre of the eyes) as defined in clause 5.6.4. Figure 18 shows a sample 2D image with the highlighted inner region.

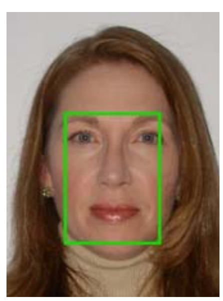


Figure 18 – Sample 2D image of the inner region.

12.6 Requirements on Full Frontal 3D Image Types using the 3D Point Map Representation

12.6.1 3D Point Map Width and Height

The resolution of the 3D Point Map is directly dependent on the Width and Height of the 3D Point Map. To enable the interchange of high resolution 3D data the minimum dimensions for Full Frontal 3D Image Types using the 3D Point Map representation shall be

$3DPointMapWidth_{\min} = 140$ Pixels and

$3DPointMapHeight_{\min} = 170$ Pixels.

12.6.2 Face Coverage

For biometric purposes it is eminently important that the outer region, as defined in clause 12.5.2 shall be covered by sufficient measurement points. Therefore, for a Full Frontal 3D Image Type using the 3D Point Map a minimum of 70% of the points shall have X and Y coordinates with -1,75 w ≤ X ≤ 1,75 w and -1,75 w ≤ Y ≤ 2,55 w in the Cartesian coordinate system with the Landmark Point Prn as origin, where w is the distance between the feature points 12.1 and 12.2 (centre of the eyes) as defined in clause 5.6.5.

12.7 Requirements on Full Frontal 3D Image Types using the 3D Vertex Representation

12.7.1 Face Coverage

For biometric purposes it is eminently important that the inner region, as defined in clause 12.5.3, shall be covered by sufficient measurement points. Therefore, for a Full Frontal 3D Image Type using the Vertex representation a minimum of 1000 points shall have X and Y coordinates with -1,5 w ≤ X ≤ 1,5 w and -1,8 w ≤ Y ≤ 1,8 w in the Cartesian coordinate system with the Landmark Point Prn as origin, where w is the distance between the feature points 12.1 and 12.2 (centre of the eyes) as defined in clause 5.6.4. There shall be at least one vertex point projected on the plane of the inner region per square centimetre with coverage of 80% in the inner region.

13 The Token Frontal 3D Image Type

The Token Frontal 3D Image Type shall fulfil the following requirements.

13.1 Inheritance requirements

The Token Frontal 3D Image Type inherits all requirements of the Basic Image Type, the requirements of the Full Frontal 3D Image Type as defined in clauses 12.2 to 12.4. Furthermore, it inherits all requirements of the Token Frontal Image Type.

13.2 Requirements on Token Frontal 3D Image Types using the Range Image Representation

The Token Frontal 3D Image Type using the Range Image representation inherits all requirements of the Full Frontal 3D Image Type using the Range Image representation as defined in clause 12.5.

13.3 Requirements on Token Frontal 3D Image Types using the 3D Point Map Representation

The Token Frontal 3D Image Type using the 3D Point Map representation inherits all requirements of the Full Frontal 3D Image Type using the 3D Point Map representation as defined in clause 12.6.

13.4 Requirements on Token Frontal 3D Image Types using the Vertex Representation

The Token Frontal 3D Image Type using the using the Vertex representation inherits all requirements of the Full Frontal 3D Image Type using the using the Vertex representation as defined in clause 12.7.

Annex A
(informative)

Best practices for Face Images

A.1 Basic Face Images

A.1.1 Purpose

This clause discusses specifications acknowledged to be important to the fulfilment of the stated purposes of record creation and yet too stringent or ill defined to appear in the normative clauses of this document. It should be read in conjunction with clause 6 "The Basic Face Image Type”.

A.1.2 Landmark Point determination

The Landmark Point block defined in clause 5.6 can be added to the record format of any Basic Face Image Type or subtype to describe the position of Landmark Points (landmarks) used by face recognition algorithms.

If possible, Landmark Points should be determined on images before compression is applied.

Landmark Points should be included in the record format if they have been accurately determined, thereby providing the option that these parameters do not have to be re-determined when the image is processed for face recognition tasks.

Typically a computer algorithm will either accurately determine the position of the Landmark Point or completely fail and provide either clearly erroneous or no landmark information. Therefore, a method for accurate determination is the use of computer-automated Landmark Point determination followed by human verification and potential override of the computer determined Landmark Points.

A.1.3 Recommended Landmark Points

For the proper registration of the face in all dimensions, a minimum of three Landmark Points that span a surface is needed. For the sake of accuracy more points should be determined, and these points should be distributed over the entire face. To be more specific, it is recommended to add the following Landmark Points to any Facial Record (ref Figure 7 and Figure 8): the middle point of the eyes (12.1 and 12.2), the base of the nose (9.4, 9.5, and 9.15) and the upper lip of the mouth (8.4, 8.1 and 8.3).

A.2 Frontal Images

A.2.1 Purpose

This clause discusses specifications acknowledged to be important to the fulfilment of the stated purposes of Frontal Image capture and creation yet too stringent or ill defined to appear in the normative clauses of this document. It should be read in conjunction with clause 7 "The Frontal Image Type”.

A.2.2 Assistance in positioning the face

Hands, arms etc. of an assisting person used to support the positioning should not be visible.

A.2.3 Pose

The full-face frontal pose should be used. Rotation of the head should be less than +/- 5 degrees from frontal in every direction – roll, pitch and yaw (ref. 5.5.8).

A.2.4 Expression

The expression should be neutral (non-smiling) with both eyes open normally (i.e. not wide-open), and mouth closed (mouth is closed if the distance between Landmark points 2.2 and 2.3 is less than 50% of the distance of landmark points 2.3 and 8.2 from Figure 7). Every effort should be made to have supplied images comply with this specification. A smile with closed jaw is not recommended. Examples of non-recommended expressions are

a) Smile,

b) Raised eyebrows,

c) Eyes looking away from the camera,

d) Squinting,

e) Frowning.

A.2.5 Examples of unacceptable clarity of eyes/eye sockets

a) Closed eyes. Eyes are considered open if iris and pupil are visible.

b) Hair covering eyes. Hair shall not cover any part of the eyes. It is recommended that hair should not cover Landmark Points 3.2, 3.8, 3.12 for the right eye and Landmark Points 3.1, 3.7 and 3.11 for the left eye in Figure 7, as well as region above these points that measures 5% of inter-eye distance. An example of unacceptable image is shown in Figure 19.

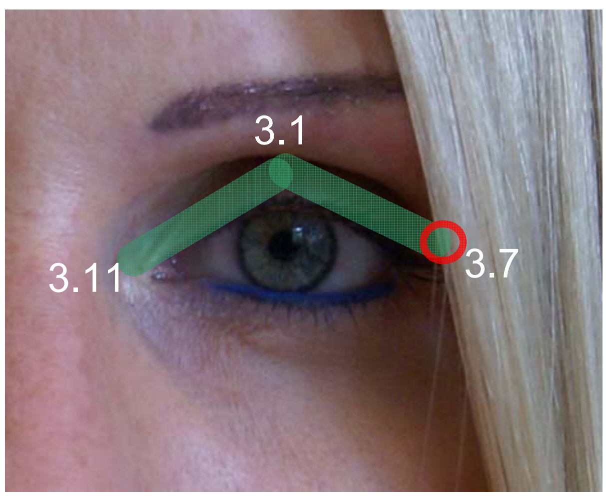


Figure 19 — Example of an unacceptable image due to hair covering the eyes

c) Rim of glasses covering part of the eye. Rims of glasses are covering part of the eye if any part of rims covers Landmark Points 3.2, 3.4, 3.8 and 3.12 for the right eye and Landmark Points 3.1, 3.3, 3.7 and 3.11 for the left eye in Figure 7, as well as region around these points that measures 5% of inter-eye distance. If rims of glasses are not visible or are completely transparent, they cannot cover any part of the eye. An example of unacceptable image is shown in Figure 20.

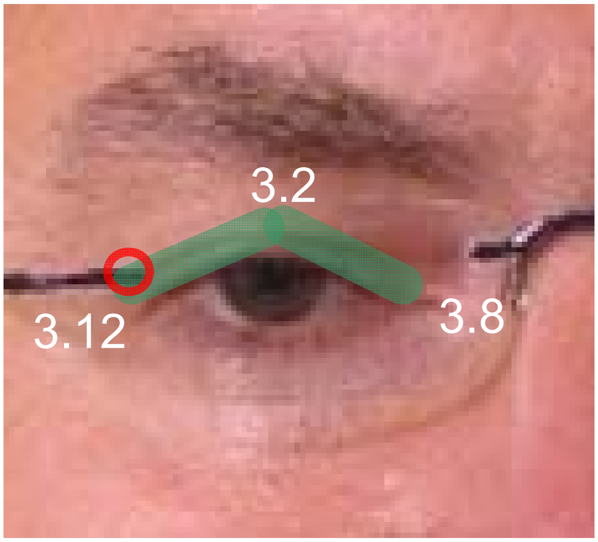


Figure 20 — Example of an unacceptable image due to rims covering the eyes

A.2.6 Eye glasses

If the person normally wears glasses then they should wear glasses when their photograph is taken.

If glasses are worn that tint automatically under illumination, they should be photographed without tint by tuning the direct illumination or background lighting. Only in abnormal cases where the tint cannot be reduced should the glasses be removed. In cases where tinted glasses are worn, the specification of dark glasses in the header structure is recommended.

Lighting artefacts can typically be avoided by increasing the angle between the lighting, subject and camera to 45º (degrees) or more.

A.2.7 Head coverings

It is important that head coverings and shadows should be absent. An exception applies to cases in which a subject cannot remove a headdress, veil or scarf (e.g. for religious reasons) - in such cases the capture process should minimize shadows and obscuration of the facial features in the facial region. This might involve adjustment of the head coverings.

A.2.8 Shadows in eye-sockets

There should be no shadows in the eye-sockets due to the brow. The iris and pupil of the eyes should be visible.

NOTE This requirement is intended to exclude images in which the eyes are closed (e.g. during a blink) or half closed.

A.2.9 Backgrounds

The discussion of background is important for computer face recognition because the first step in the computer face recognition process is the segmentation of the face from the background for the purpose of registration (landmark determination). In this context, certain common problems should to be avoided if possible.

A.2.9.1 Background segmentation

The boundary between the head and the background should be clearly identifiable about the entire subject (very large volume hair excepted).

A.2.9.2 Background shadows

There should be no shadows visible on the background behind the face image.

A.2.9.3 Background uniformity

The background should be plain, and shall contain no texture containing lines or curves that could cause computer face finding algorithms to become confused. Therefore the background should be a uniform colour or a single colour pattern with gradual changes from light to dark luminosity in a single direction.

A.2.9.4 Background examples

A typical background to enhance machine-assisted face recognition performance is 18% grey with a plain smooth surface. Plain light coloured backgrounds such as light blue are also acceptable. A white background is acceptable provided there is sufficient distinction between the face/hair area and the background.

A.2.10 Focus and depth of field

In a typical photographic situation, for optimum quality of the captured face, the f-stop of the lens should be set at two (or more) f-stops below the maximum aperture opening when possible to obtain enough depth of field.

Greater than one millimetre spatial sampling rate will be considered accomplished if the individual millimetre markings of rulers placed on the subject's nose and ear facing the camera can be seen simultaneously in a captured test image.

If the camera lacks auto focus all subject positions will need to be maintained in a defined area for all image captures.

A.2.11 Greyscale density

The dynamic range of the image should have at least 7 bits of intensity variation (span a range of at least 128 unique values) in the facial region of the image. The facial region is defined as the region from crown to chin and from the left ear to the right ear. This recommendation may require camera, video digitizer, or scanner settings to be changed on an individual basis when the skin tone is excessively lighter or darker than the average (preset) population.

A.2.12 Colour saturation

The colour saturation of a 24-bit colour image should be such that after conversion to greyscale, there are 7 bits of intensity variation in the facial region of the image.

A.2.13 No unnatural colour

Greyscale photographs should be produced from common incandescent light sources. Colour photographs should use colour-balancing techniques such as using high colour-temperature flash with standard film or tungsten-balanced film with incandescent lighting.

A.2.14 Colour calibration

Colour calibration using an 18% grey background or other method (such as white balancing) is recommended.

A.2.15 Distortion of the camera lens

The purpose of this requirement is to make consistent radial distortion due to focal length. For a typical photo capture system with a subject 1,5 to 2,5 meters from the camera, the focal length of the camera lens should be that of a medium telephoto lens. For 35 mm photography this means that the focal length should be between 90 mm and 130 mm. For other negative formats/sensors the recommended focal length is 2 to 3 times the diagonal of the negative/sensor.

Figure 21 shows an example, of how the distance to subject and the selection of the camera lens influence the image captured.

Figure 21 – Samples of correct (left) and incorrect (middle, right) selection of focal length for a given distance to subject

A.3 Full Frontal Images

A.3.1 Digital attributes of Full Frontal Images

A.3.1.1 Photo Spatial Sampling Rate

For an image for optimal human examination and permanent storage, the preferred minimum spatial sampling of the full image is at least 240 pixels of spatial samples for the width of the head, and correspondingly roughly 120 pixels from eye centre to eye centre. This corresponds to a minimum full image width of 420 pixels and an image height of 525 pixels.

1) For a photograph with head width 20mm (roughly 0,78 inches), the recommended scanner spatial sampling rate is 120 dots per centimetre (roughly 300 dots per inch).

2) For a photograph with head width 13 mm (roughly 0,5 inches), the recommended scanner spatial sampling rate is 189 dots per centimetre (roughly 480 dots per inch).

3) For a photograph with head height (from chin to crown) of 25mm (roughly 1 inch), this in turn corresponds to a head width on average of roughly 20 mm (roughly 0,8 inches) using a typical head geometric ratio of 4 to 5. This corresponds to a required scanner spatial sampling rate of 117 dots per centimetre (roughly 300 dots per inch).

Therefore when colour scanning supplied paper photograph portraits of conforming dimensions using a scanner, the colour scanner spatial sampling rate should typically be set to 300dpi.

On the other hand, if photographs are scanned at 120 pixels per centimetre (300ppi) the requirement of 90 pixels minimum inter-eye distance corresponds to an inter-eye distance of approximately 8mm. In analogy, the best practice requirement of 120 pixels minimum inter-eye distance corresponds to an inter-eye distance of approximately 10mm for photographs scanned at 120 pixels per centimetre (300ppi).

A.3.1.2 Post processing

Multiple compression should be avoided to generate Full Frontal Images.

A.3.2 Best practices for use of Full Frontal Images on travel documents

A.3.2.1 Subject Height field

The subject height field as defined in clause 5.5.5 should be specified.

A.3.2.2 Width to height ratio of the image

For a Full Frontal Image, the (Image Width: Image Height) aspect ratio should be between 1:1,25 and 1:1,34.

This allows for ratio of 1:1,25 specified by NIST best practices for mug shots, 1:1,28 used in many passport images, and 1:1,33 used in many driver's license images.

A.3.2.3 Head size relative to the image size

For a Full Frontal Image the (Image Width: Head Width) ratio (A:CC) should be between 7:5 and 2:1 as this satisfies requirements from numerous driver's license and international passport agencies.

For cases where the subject has a lot of hair, this constraint is more important than including the entire hairline in the photograph.

For teens and adults, the crown to chin portion of the full-face frontal pose should occupy 70% to 80% of the vertical length of the image as this satisfies requirements from numerous driver's license and international passport agencies.

For children, typically defined as persons under the birth age of 11 years, a smaller head size of 50% of the image area is acceptable if required to maintain photographic quality of the image such as to avoid distortion such as fish eye (ref. 4.11) or blurring.

A.3.2.4 Summary of best practice photographic recommendations

For convenience Table 39 summarizes the geometric and pose constraints in clauses A.2.3 and A.3.2.2 - A.3.2.3.

Table 39 — Summary best practices for Full Frontal Images on travel documents

Clause	Definition	Recommendation
A.2.3	Pose	± 5 degrees from frontal in roll, pitch and yaw
A.3.2.2	Width to Height Ratio of Image	$1{,}25 \le B/A \le 1{,}34$
A.3.2.3	Width of Head	$1{,}4\ CC \le A \le 2\ CC$
A.3.2.3	Length of Head	$0{,}7\ B \le DD \le 0{,}8\ B$
A.3.2.3	Length of Head (Children under the age of 11)	$0{,}5\ B \le DD \le 0{,}8\ B$

Sample images and sample photograph taking guidelines for travel documents

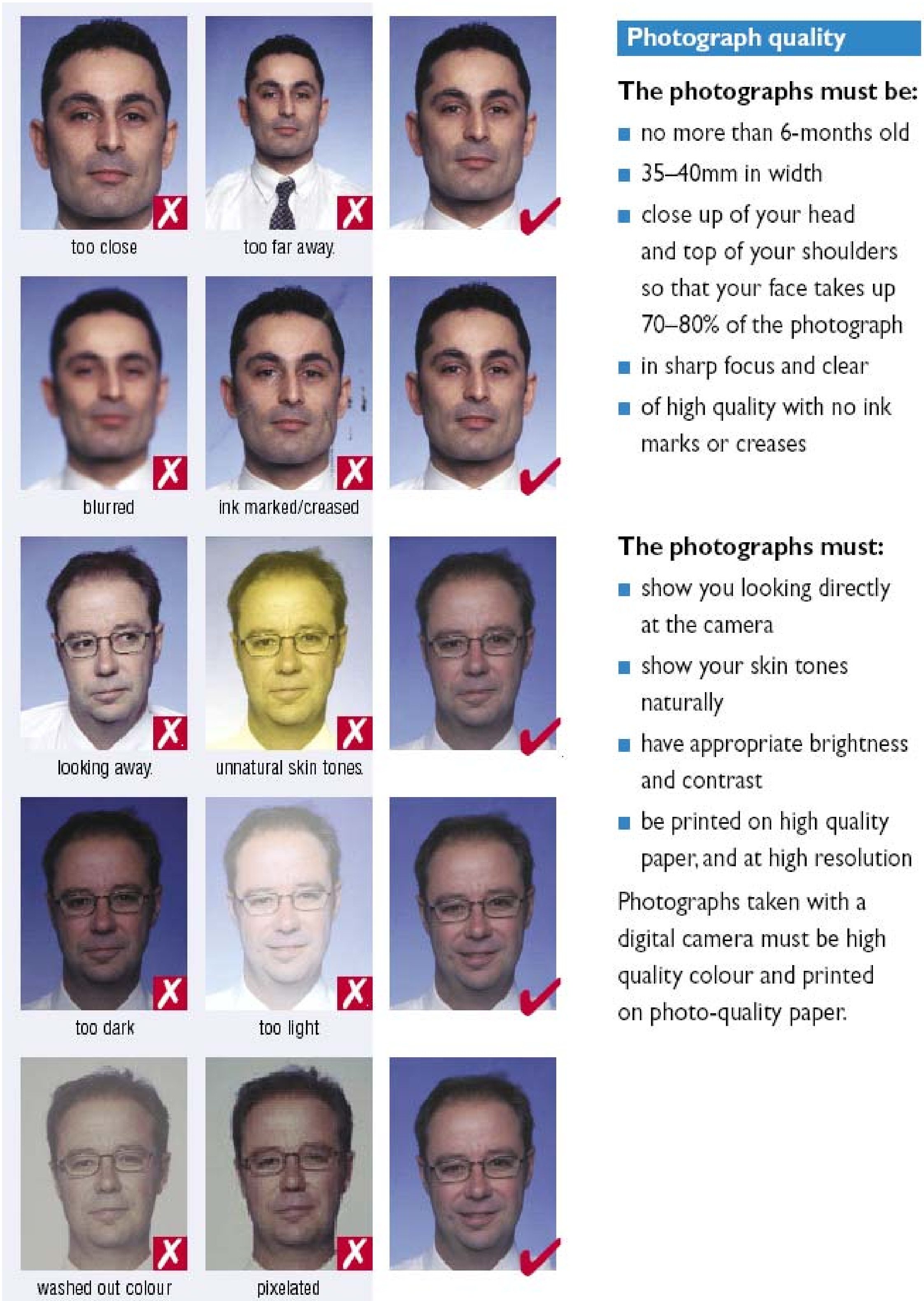


Figure 22 (45 images) — Best practices for the purpose of travel document creation

Style and lighting

The photographs must:

- be colour neutral
- show your eyes open and clearly visible—no hair across your eyes
- show you facing square on to the camera, not looking over one shoulder (portrait style) or tilted, and showing both edges of your face clearly
- be taken with a plain light-coloured background
- be taken with uniform lighting and not show shadows or flash reflections on your face and no red eye

Figure 22 (45 images) — Best practices for the purpose of travel document creation

Glasses and head covers

If you wear glasses:

- the photograph must show your eyes clearly with no flash reflection off the glasses, and no tinted lenses (if possible, avoid heavy frames— wear lighter framed glasses if you have them)
- make sure that the frames do not cover any part of your eyes.

Head coverings:

- are not permitted except for religious reasons, but your facial features from bottom of chin to top of forehead and both edges of your face must be clearly shown.

Expression and frame

Your photographs must:

- show you alone (no chair backs, toys or other people visible), looking at the camera with a neutral expression and your mouth closed.

Figure 22 (45 images) — Best practices for the purpose of travel document creation

A.3.3 Full Frontal Image Compression

A.3.3.1 Compression — no region of interest

Face recognition performance results for the compression of Full Frontal Images are shown in Figure 23 and Figure 24 below, from faces obtained from Passports Australia within the Australian Department of Foreign Affairs and Trade[3]. Here, 1 000 matching pairs (original and renewals) of real passport images were considered.

These images were originally scanned at 300 dpi and have standard passport photo size geometric characteristics of width and height of 416 x 536 pixels, with the head dimensions corresponding to clause 8. The average size of the original, uncompressed images was approximately 669 KB. The images used in these tests were compressed to an average size of 71 KB using JPEG, then decompressed and recompressed using JPEG and JPEG2000 for the matching tests.

This initial compression of the images could potentially cause JPEG artefacts to be present in the images used in the tests, but given the relatively low compression ratio of 10:1 used, this initial compression has shown to have had a negligible impact on the outcome of the data as it pertains to matching performance.

The face registration and face recognition technologies used for this analysis were FaceIt version 5.0, from Identix Corporation, and ZN-FaceRecServer version 1.1, from ZN Vision Technologies AG. The images were computer aligned (eye positions were determined by computer).

A set of full-face images was compressed and matched against a set of uncompressed full images. The correct match probability rank one statistic was studied as a function of compression level. The rank one statistic denotes the number of times the top match was the correct match in a one-to-many search attempt (where a correct match is always possible). It is a function of the size of the database, and this statistic is commonly used in the context of facial one-to-many searches.

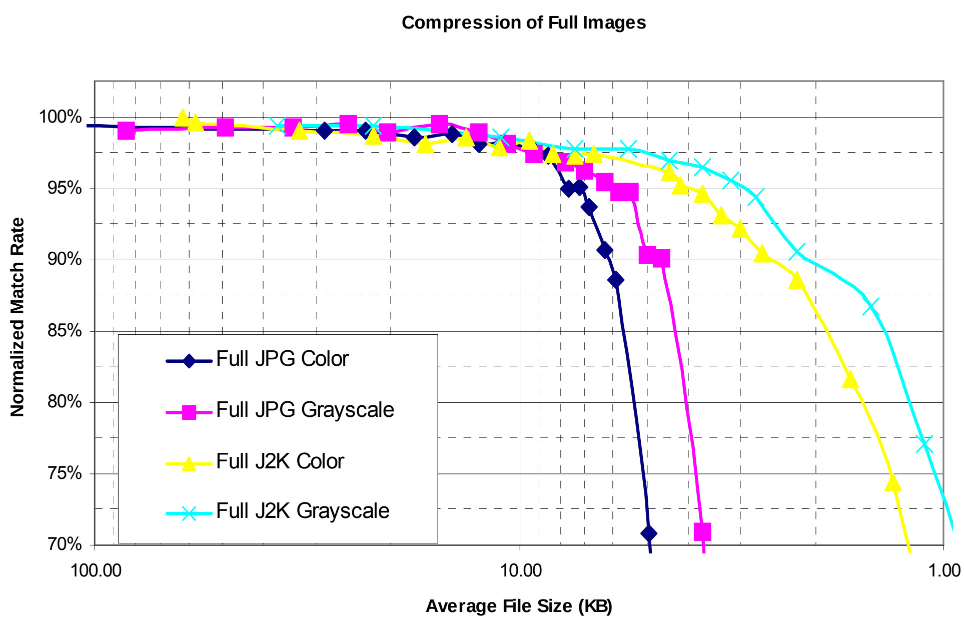


Figure 23 — Identix, compression of Full Frontal Images versus face recognition performance

3 Provided by Terry Hartmann, Passports Australia, September 2002.

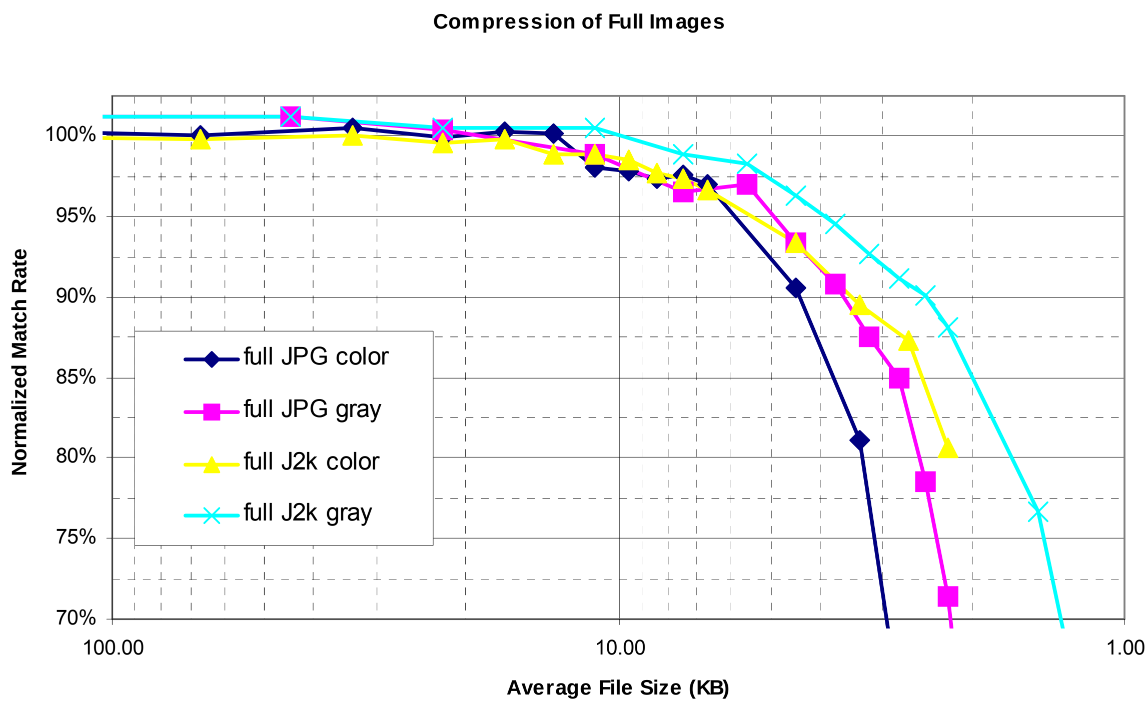


Figure 24 — ZN Vision Technologies, compression of Full Frontal Images versus face recognition performance

In Figure 23 and Figure 24, the "Normalized Match Rate" denotes the correct matches at each compression factor (each file size level), divided by the correct matches at zero compression, and then multiplied by 100 to obtain a percentage. That is, the value of 100% implies that the compression had no effect on the matching ability of the technology. A value of 50% means that only 50% of the correct matches in the 1 000 one-to-many search attempts were maintained at that compression level. The "Average File Size" denotes the compressed file size.

The key factor of interest in this analysis is how much a facial image can be compressed before matching performance degrades significantly (by more than 1-2%) over the results achieved with no compression.

The graph in and show that:

1. The performance degrades quickly below a compressed file size of 10KB.

2. JPEG2000 performs relatively better than JPEG in this analysis.

A.3.3.2 Recommendations for maximum compression and file sizes for JPEG and JPEG2000

For the purpose of making recommendations, a significant degradation has been defined as greater than 2%, hence these represent the minimum file sizes and compression rations to achieve no more than 2% degradation when compressing images compared with the results achieved with no (or very minimal) compression. Results have been rounded to the nearest 1K.

In conclusion, for use of these two technologies for automatic face one-to-many searches, the compressed image file size should be no lower than 11 KB on average for Full Frontal images similar to those used in the experiment (passport images).

A.3.4 Full Frontal Image compression using region of interest

A.3.4.1 Discussion

A Full Frontal or Token Frontal image can be compressed further in situations where the alignment of the eyes is known precisely, either by use of a well-studied eye location algorithm, or by human verification of eye positions.

JPEG2000 can be used to implement "region of interest" (ROI) compression, as it is a technique specified in the ISO JPEG2000 standard and well defined for JPEG2000 software libraries.

JPEG2000 ROI encoding can be used to achieve smaller file sizes. The Inner Region of a facial image used for matching can be compressed to a low ratio, while the Outer Region of the image is compressed to a higher ratio. The resulting image is smaller in size, but those parts of the image used for matching retain high quality while the remainder of the image maintains their usefulness for visual inspection. A standard compliant JPEG2000 decoder with ROI support will decode an ROI image regardless of the location of ROI regions.

The use of "region of interest" compression for situations where computer alignment is performed without human verification is not recommended.

A.3.4.2 Inner and outer regions, Full Image

It is important to note that additional compression can be achieved by defining inner and outer regions that are based on the face area.

Example

For example, when derived from a 300 dpi Full Frontal image, an inner region can be defined as including the entire face from crown to chin and ear to ear.

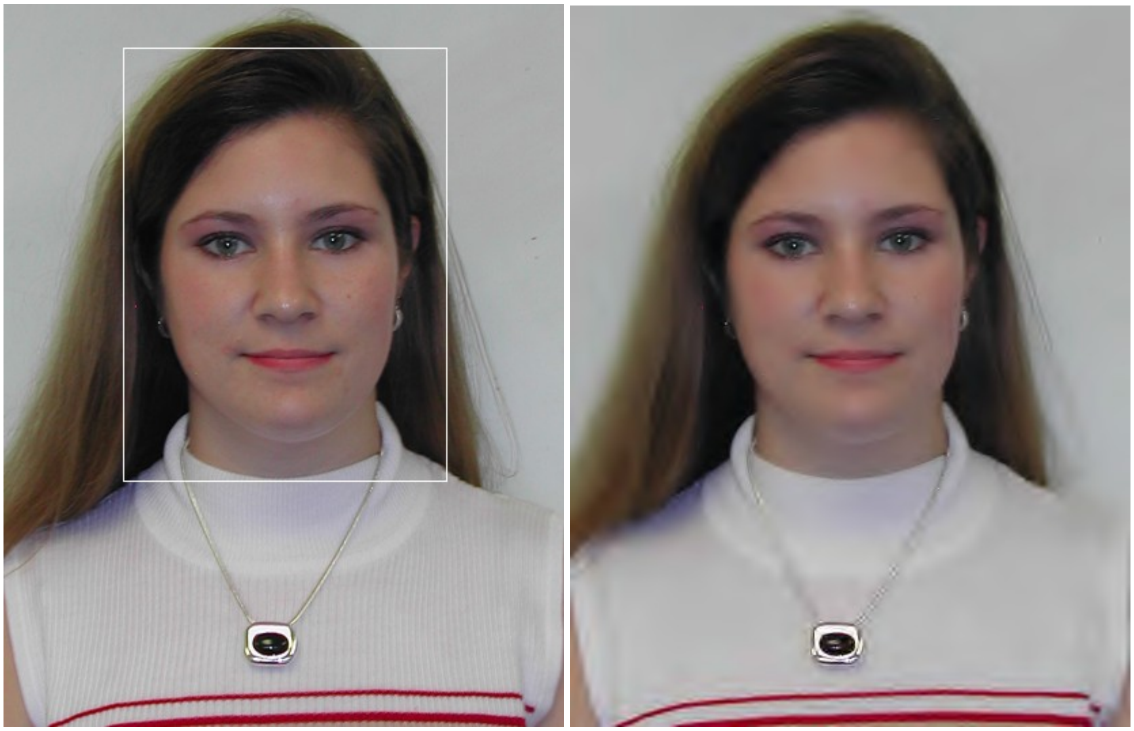
Analysis above indicates that a compression ratio of 60:1 using JPEG2000 preserves matching performance. If a 50:1 ratio is used for the Inner Region, and 200:1 can be used on the Outer Region with an acceptable level of degradation for visual inspection purposes. For a colour, 300 dpi, 35x45mm JPEG2000 image (413x531 pixels, 658 KB uncompressed), with a 240x320 (230,4 KB) Inner Region as defined in Figure 25,

A. 200:1 Outer Region: (658-230,4 KB)/200 = 2,14 KB

B. 50:1 Inner Region: (230,4 KB)/50 = 4,61 KB

Total file size: 2,14 + 4,61 = 6,75 KB. File size reduction: ~40%

In Figure 25 below, the image on the left represents the uncompressed image, and shows the bounds of the Inner Region. The image on the right is compressed using JPEG2000 ROI as described above.



a) b)

Figure 25 — Example uncompressed (a) and compressed (b) using region of interest shown in (a)

A.4 Token Images

A.4.1 Token image sizes

As discussed in clause 9, the Width variable of the Token image defines the geometry of the face using eye position landmarks. The minimum width is 240 pixels, which corresponds to an inter-eye distance of 60 pixels inclusive. There is no maximum.

Interpolation required in the affine transformation used to create a Token image can have the effect of introducing artefacts that can harm the face recognition process. For example, if one company chose to use 70 pixels from eye to eye while another chose 60, there might be unnecessary problems. Therefore, in order to improve the interoperability of Token images, it is recommended that the width be specified in units of 240. Examples are given in Table 40.

Table 40 — The recommended width variables for use with Token Images

Width	Distance from Eye to Eye (Inclusive)
240	60
480	120
720	180

A.4.2 Creation of a Token Image

Figure 26 depicts an example of the steps that can be involved in the transformation of an image to a Token Face image. In the creation of a 240-pixel wide Token Face image, the original image (a) is rotated to horizontally align the eyes (b). The image is then uniformly scaled so that there are exactly 60 pixels between the centres of the eyes (c). Lastly the image is translated and cropped (d) such that the first eye coordinate is (89,144) i.e. 89 pixels over and 144 pixels down from the upper left corner of the image (0,0). The black pixels which are padding the borders can be any colour.

Figure 26 — Affine transformation and cropping

A.4.3 Best practices for digital attributes of Token Images

The use of a computer algorithm with additional human visual inspection of the computer-generated results is the recommended method for determination of eye positions.

The normative practice is to fill the area with any colour.

A bi-linear or other advanced interpolation and sampling method is recommended in the scaling and rotation stages of the affine transformation.

A.4.3.1 Post processing

Multiple compression should be avoided to generate Token Frontal Images.

A.4.4 Token Image compression

A.4.4.1 Compression — no region of interest

Face recognition performance results for the compression of Token images are shown in Figure 27 and Figure 28 below, from faces obtained from Passports Australia within the Australian Department of Foreign Affairs and Trade[4]. Here, 1 000 matching pairs (original and renewals) of real passport images were considered.

See clause A.3.3 for a detailed description of the compression experiments. The same experiment for Full Frontal images discussed in clause A.3.3 was repeated for Token images. The results for Identix and ZN Vision Technologies are shown in Figure 27 and Figure 28.

[4] Provided by Terry Hartmann, Passports Australia, September 2002.

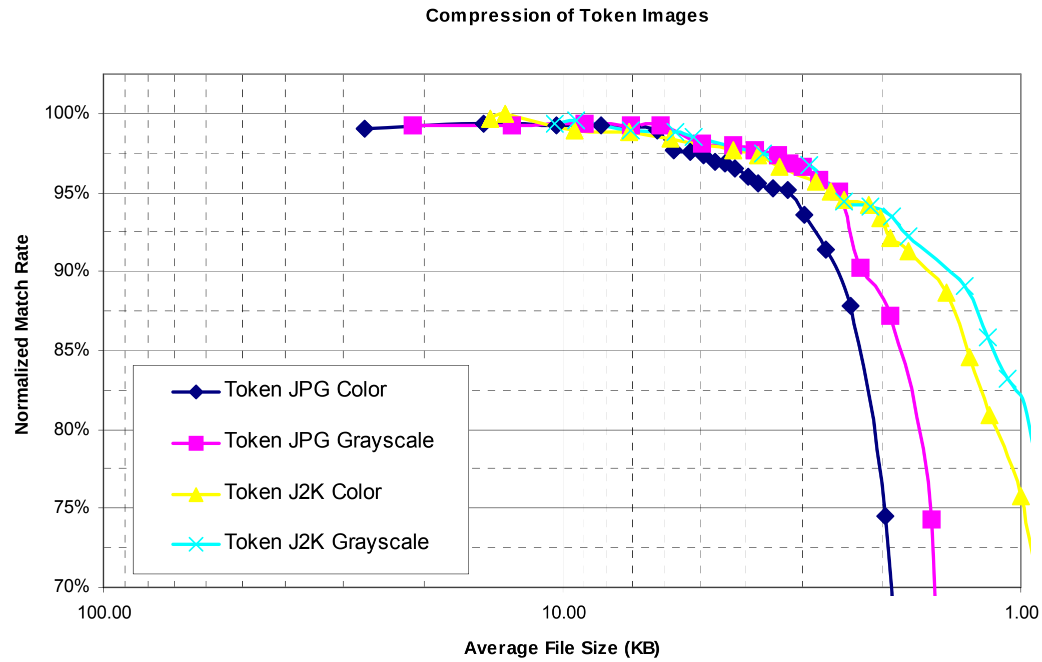


Figure 27 — Identix, compression Token images versus face recognition performance

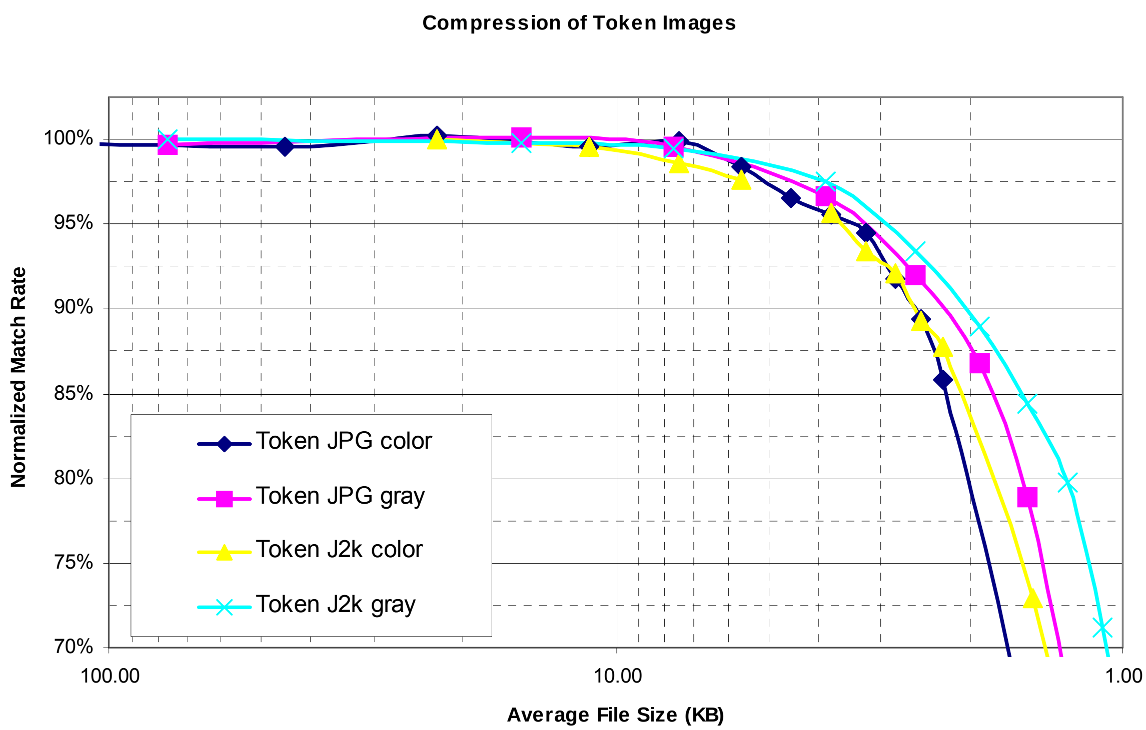


Figure 28 — ZN Vision Technologies, compression of Token Images versus face recognition performance

The graphs in Figure 27 and Figure 28 show that:

1. The performance degrades quickly below a compressed file size of 8KB.
2. JPEG2000 performs relatively better than JPEG in this analysis.

A.4.4.2 Recommendations for maximum compression and file sizes for JPEG and JPEG2000 Token images

For the purpose of making recommendations, a significant degradation has been defined as greater than 2%, hence these represent the minimum file sizes and compression ratios to achieve no more than 2% degradation when compressing images compared with the results achieved with no (or very minimal) compression. Results have been rounded to the nearest 1K.

In conclusion, for use of these technologies for automatic face one-to-many searches, the compressed Token image file size should be no lower than 9 KB on average, with JPEG or JPEG2000.

A.4.5 Token Image compression using region of interest

A.4.5.1 Discussion

A Full Frontal or Token Frontal image can be compressed further in situations where the alignment of the eyes is known precisely, either by use of a well-studied eye location algorithm, or by human verification of eye positions.

JPEG2000 can be used to implement "region of interest" (ROI) compression, as it is a technique specified in the ISO JPEG2000 standard and well defined for JPEG2000 software libraries.

JPEG2000 ROI encoding can be used to achieve smaller file sizes. The Inner Region of a facial image used for matching can be compressed to a low ratio, while the Outer Region of the image is compressed to a higher ratio. The resulting image is smaller in size, but those parts of the image used for matching retain high quality while the remainder of the image maintains their usefulness for visual inspection. A standard compliant JPEG2000 decoder with ROI support will decode an ROI image regardless of the location of ROI regions.

The use of "region of interest" compression for situations where computer alignment is performed without human verification is not recommended.

A.4.6 Inner and outer regions for the Token Image for the purpose of compression

We define an Inner Region, when derived from a 240 width Token image, as a rectangular area within and including the pixel positions (24, 24), (215, 24), (24, 263), and (215, 263) as shown Figure 29. This is a region of 192 x 240 pixels in size.

Figure 29 shows the dimensions of the inner region when width W = 240. The generalized coordinates for the Inner Region are (0,1*W, 0,1*W), (0,9*W-1, 0,1*W), (0,1*W , 1,1*W-1), and (0,9*W-1, 1,1*W-1).

The Outer Region is the entire full image region excluding the Inner Region, and is of area 240x320-192x240 = 30 720 pixels, or (75-45) x 3 KB = 90 KB of size in bytes for 3 bytes per pixel.

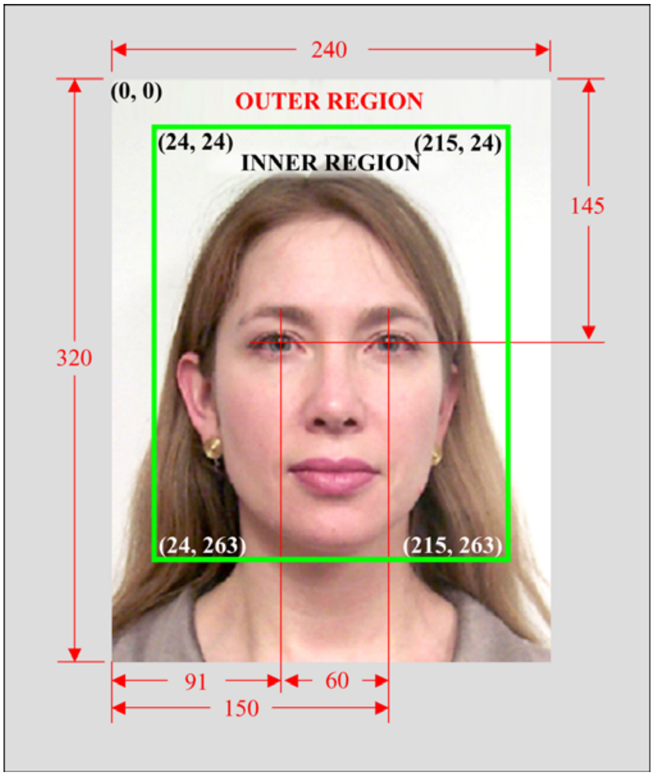


Figure 29 — Suggested region of interest for Token Images

Compression versus performance results for Token images with region of interest as defined above are not provided here.

A.5 Best Practices for the Full Frontal 3D Image Type

Besides the requirements of clause 12 it is best practice for a Full Frontal 3D Image to fulfil the following requirements.

A.5.1 Best Practices for the 2D part of the Full Frontal 3D Image Type

The Best Practices for the Full Frontal Image type should be fulfilled.

A.5.2 Compatibility considerations

There are many implementations deployed in the field that comply to the base standard only. Therefore, for application areas where interoperability is most important, it is best practice to store the 2D part of the Full Frontal 3D image as a Full Frontal Image in the same Face Record. This allows the mentioned implementations to safely read the Full Frontal Image record and therefore interoperability is improved.

A.5.3 Pose of the 3D representation

The rotation of the head should be less than +/- 5 degrees from frontal in roll (ref. Clause 5.5.8.3).

A.5.4 3D to 2D Image Temporal Synchronicity

The 3D to 2D Image Temporal Synchronicity field should be specified and should not be filled with 0xFFFF (unspecified).

A.5.5 3D Acquisition Time

The 3D Acquisition Time field should be specified, i.e. the field should not be filled with 0xFFFF (unspecified).

A.5.6 Best Practices for Full Frontal 3D Image Types using the Range Image Representation

A.5.6.1 ScaleX, ScaleY and ScaleZ

The resolution of the stored depth data strongly depends on ScaleZ. To preserve quality a maximum value of $ScaleZ_{\max} = 0{,}8mm$ should be defined for the Full Frontal Range Image Type.

For the same reason a maximum value of $ScaleX_{\max} = ScaleY_{\max} = 0{,}8mm$ should be defined for the Full Frontal 3D Image Type using the range image representation in a Cartesian coordinate system.

A.5.6.2 Non-valid points in Range Image

At maximum 20% of the pixel of the Range Image in the region defined in 11.5 should have a zero value, indicating a non-valid depth value. Furthermore, in the "inner region" defined as [-1,5 *w*, 1,5 *w*] x [-1,8 *w*, 1,8 *w*] in the Cartesian coordinate system with the Landmark Point Prn as origin at maximum 10% of the pixel should have a zero value, indicating a non-valid depth value. Here, *w* is the distance between the feature points 12.1 and 12.2 (centre of the eyes) as defined in clause 5.6.5.

A.5.7 Best Practices for the Full Frontal 3D Image Types using the 3D Point Map Representation

A.5.7.1 3D Point Map Width and Height

As a best practice requirement for 3D Image Types using the 3D Point Map representation the minimal dimensions of the 3D Point Map should be

$3DPointMapWidth_{\min} = 175$ Pixels and

$3DPointMapHeight_{\min} = 213$ Pixels.

A.5.7.2 Face Coverage

As a best practice requirement for 3D Image Types using the 3D Point Map representation a minimum of 90% of the points should have X and Y coordinates with -1,75 *w* ≤ X ≤ 1,75 *w* and -1,75 *w* ≤ Y ≤ 2,55 *w* in the Cartesian Coordinate System with landmark Prn as origin, where *w* is the distance between the feature points 12.1 and 12.2 (centre of the eyes) as defined in clause 5.6.5.

A.5.8 Best Practices for Full Frontal 3D Image Types using the 3D Vertex Representation

A.5.8.1 Face Coverage

As a best practice requirement for 3D Image Types using the Vertex Representation a minimum of 1 500 points should have X and Y coordinates with -1,5 *w* ≤ X ≤ 1,5 *w* and -1,8 *w* ≤ Y ≤ 1,8 *w* in the Cartesian coordinate system with landmark Prn as origin, where *w* is the distance between the feature points 12.1 and 12.2 (centre of the eyes) as defined in clause 5.6.5. There should be at least one vertex point projected on the plane of the inner region per square centimetre with coverage of 90% in the inner region.

A.6 Best Practices for Token Frontal 3D Images

Besides the requirements of clause 13 it is best practice for a Frontal 3D Image associated with a Token Frontal image to fulfil the following requirements.

A.6.1 Best Practices for the 2D part of the Token Frontal 3D Image

The Best Practices for the Token Frontal Image type should be fulfilled.

A.6.2 Compatibility considerations

There are many implementations deployed in the field that comply to the base standard only. Therefore, for application areas where interoperability is most important, it is best practice to store the 2D part of the Token

Frontal 3D image as a Token Frontal Image in the same Face Record. This allows the mentioned implementations to safely read the Token Frontal Image record and therefore interoperability is improved.

A.6.3 Pose of the 3D representation

The rotation of the head should be less than +/- 5 degrees from frontal in roll (ref. Clause 5.5.8.3).

A.6.4 3D to 2D Image Temporal Synchronicity

The 3D to 2D Image Temporal Synchronicity field should be specified and should not be filled with 0xFFFF (unspecified).

A.6.5 3D Acquisition Time

The 3D Acquisition Time field should be specified, i.e. the field should not be filled with 0xFFFF (unspecified).

A.6.6 Best Practices for Token Frontal 3D Image Types using the Range Image Representation

The best practices as outlined in clause A.5.6 should be fulfilled.

A.6.7 Best Practices for Token Frontal 3D Image Types using the 3D Point Map Image Representation

The best practices as outlined in clause A.5.7 should be fulfilled.

A.6.8 Best Practices for Token Frontal 3D Image Types using the Vertex Representation

The best practices as outlined in clause A.5.8 should be fulfilled.

A.7 Summary of mandatory and best practices for the 3D Image Types

A.7.1 Normative coordinate system and pose requirements for the Basic 3D Image Type

	Range 3D	3D Point Map	3D Vertex
Cartesian coordinate system	Supported	Supported	Supported
Cylindrical coordinate system	Supported	Not supported	Not supported
Scale and Offset values used for coordinate transformation	Supported	Not supported, fixed scaling and offset	Not supported, fixed scaling and offset
Restrictions on pose (yaw, pitch, roll)	None	None	None

A.7.2 Normative coordinate system and pose requirements for the Full Frontal 3D Image Type

	Range 3D	3D Point Map	3D Vertex
Cartesian coordinate system	Nose at origin	Nose at origin	Nose at origin
Cylindrical coordinate system	Not supported	Not supported	Not supported
Scale and Offset values used for coordinate transformation	Supported	Not supported, fixed scaling and offset	Not supported, fixed scaling and offset
Restrictions on Pose (yaw, pitch, roll) in degrees	(± 5, ± 5, ± 8)	(± 5, ± 5, ± 8)	(± 5, ± 5, ± 8)

A.7.3 Normative coordinate system and pose requirements for the Token Frontal 3D Image Type

	Range 3D	3D Point Map	3D Vertex
Cartesian coordinate system	Nose at origin	Nose at origin	Nose at origin
Cylindrical coordinate system	Not supported	Not supported	Not supported
Scale and Offset values used for coordinate transformation	Supported	Not supported, fixed scaling and offset	Not supported, fixed scaling and offset
Restrictions on pose (yaw, pitch, roll) in degrees	(± 5, ± 5, ± 8)	(± 5, ± 5, ± 8)	(± 5, ± 5, ± 8)

A.7.4 Best practice coordinate system and pose requirements for the Full Frontal 3D Image Type

	Range 3D	3D Point Map	3D Vertex
Cartesian coordinate system	Nose at origin	Nose at origin	Nose at origin
Cylindrical coordinate system	Not supported	Not supported	Not supported
Scale and Offset values used for coordinate transformation	Supported	Not supported, fixed scaling and offset	Not supported, fixed scaling and offset
Restrictions on pose (yaw, pitch, roll) in degrees	(± 5, ± 5, ± 5)	(± 5, ± 5, ± 5)	(± 5, ± 5, ± 5)

A.7.5 Best practice coordinate system and pose requirements for the Token Frontal 3D Image Type

	Range 3D	**3D Point Map**	**3D Vertex**
Cartesian coordinate system	Nose at origin	Nose at origin	Nose at origin
Cylindrical coordinate system	Not supported	Not supported	Not supported
Scale and Offset values used for coordinate transformation	Supported	Not supported, fixed scaling and offset	Not supported, fixed scaling and offset
Restrictions on pose (yaw, pitch, roll) in degrees	(± 5, ± 5, ± 5)	(± 5, ± 5, ± 5)	(± 5, ± 5, ± 5)

Annex B
(informative)

Conditions for Taking Photographs

B.1 Scope

The purpose of this annex is to provide expert guidance (i.e., best practices) for the photography of faces, especially when the resulting images are to be used for purposes of identification, either by automated face recognition systems or by human viewers. This guidance is intended for owners and operators of photography studios, photo stores and other organizations producing or requiring either conventional printed photographs or digital images of faces that may be used in applications for passports, visas, or other identification documents and when those images are required to conform to the frontal image types of this part of ISO/IEC 19794. This guidance is also intended for the designers and operators of photo booths, if those booths are required to provide face images conforming to the specifications of this standard. This annex may also be appropriate source material to application developers, application profile standard developers, or others making more general use of this standard.

There are many factors that affect face recognition system performance, including the individual’s appearance, such as his or her facial characteristics, hair style, and accessories, and the acquisition conditions, such as the camera’s field-of-view, focus, depth-of-field, background, and lighting. The acquisition conditions have, potentially, a greater influence on face recognition accuracy than the individual’s appearance and, of course, are controllable by the preparer of the face images.

This annex provides recommendations for acquiring two-dimensional (2D) face images directly with an analogue, digital, or video camera, as well as for image data acquired through traditional photo printing and digital scanning. [The acquisition of three-dimensional (3D) images is out of the scope of this annex.]

This annex may also be appropriate source material for application developers, application profile standard developers, or others making more general use of this standard.

B.2 Photography recommendations

This clause provides recommendations for photographing (acquiring) face images in a portrait studio, photo store, photo booth, registration office, or other facility. Guidance concerning the positioning of the subject and camera is provided, as well as several examples of alternative lighting arrangements. The intent of this guidance is to ensure that the subject’s face is properly positioned and uniformly illuminated, thereby producing images that are compliant with this International Standard and are without shadows or hot spots on the face or excessive glare in eyeglasses.

B.2.1 Recommendations for a photo studio or store

A photo studio or a photo store is typically a professionally operated facility, equipped with an analogue or digital camera, multiple adjustable light sources, a suitable background or backdrop cloth, and subject positioning apparatus designed to obtain high quality portraits. This section provides expert guidance for the owners and operators of such facilities when they must produce photographs compliant with the requirements of this standard. It is important Photo Studio or store operators do NOT manipulate the image in any way. It is vital photographs are not altered, as this will cause issues with applicants live comparison image capture comparisons, at customs / boarder control areas.

B.2.1.1 Recommended positioning and distance between camera and subject

The following recommendations concern the positioning of the subject and the camera.

— The camera-to-subject distance should be within the range of 1,2 to 2,5 m. Arranging the lighting without creating shadows will likely be difficult if the camera is placed any closer to the subject.

— Proper focus and depth-of-field will be assured by pre-focusing the lens at the distance of the subject's eyes and by selecting an appropriate aperture (F-stop) to ensure a depth-of-field of at least 10 centimetres, or approximately the distance from a subject's nose to ears. The depth-of-field of a lens is dependent upon its focal length, its effective aperture, and the focus distance. Point sources which are closer or farther than the distance at which a lens is well focused will be blurred, with the extent of the blur described by a "circle of confusion." If the maximum diameter of the circle of confusion is limited by, for example, the spacing between adjacent pixels in a CCD image sensor, the front and rear distances from the plane of optimum focus that produce acceptably focused images can be determined. The sum of these front and rear distances is the depth-of-field (D_{DoF}).

$$D_{DoF} = D_{front} + D_{rear}$$

$$D_{front} = \frac{cFs(s-f)}{f^2 + cF(s-f)}$$

$$D_{rear} = \frac{cFs(s-f)}{f^2 - cF(s-f)}$$

where :

D_{front} = the front focal distance, the distance from the plane of focus to the plane closest to the lens that is still in acceptable focus,

D_{rear} = the rear focal distance, the distance from the plane of focus to the plane farthest from the lens that is still in acceptable focus,

c = the diameter of the circle of confusion,

s = the distance from the lens to the object plane (subject's face), and

$F = f/a$ is the F-stop, the lens focal length f divided by the effective lens aperture a

Figure 30 illustrates these dimensions.

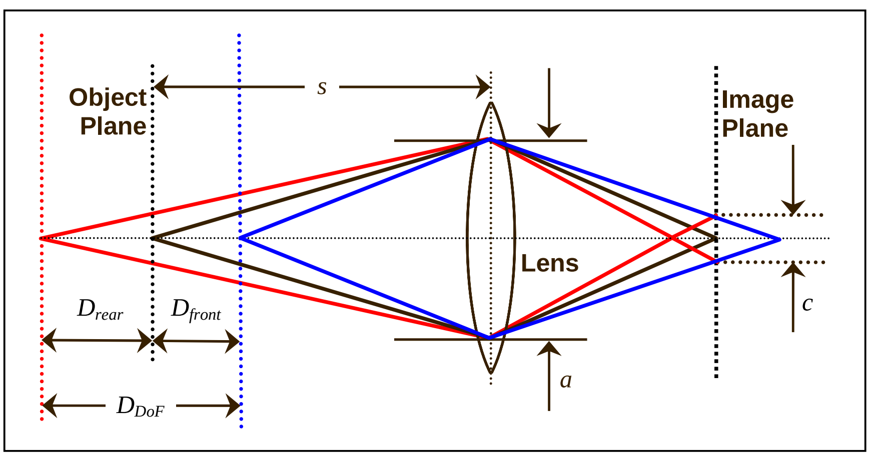


Figure 30 — Dimensions for depth-of-field calculations

- The optimum height of the camera is at the subject's eye-level. Height adjustment can be done by either using a height-adjustable stool or adjusting the tripod's height.
- The subject should be instructed to look directly at the camera and to keep his or her head erect and shoulders square to the camera. The rotation of the head should conform to the requirements of 7.2.2.

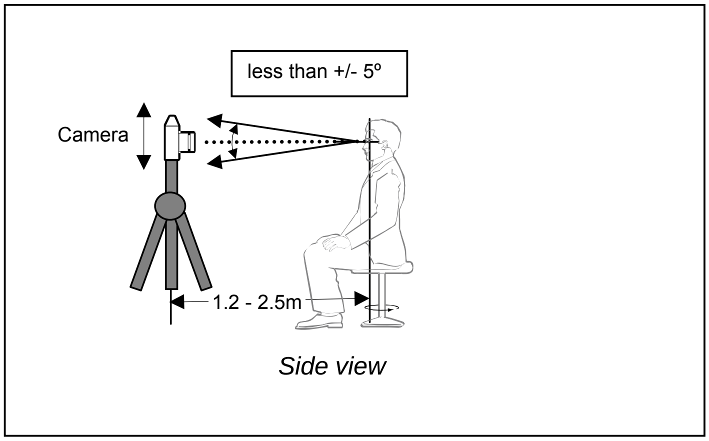


Figure 31 — Preferred distance and alignment of camera and subject

B.2.1.2 Example of exposure metering at various spots on a subject

The figure below illustrates exposure value (EV) measurement at four spots on a subject's face, namely the left and right cheeks, forehead, and chin. The measurements may be made by placing an incident light meter at the position of a subject's face and pointing the meter towards the camera. The four readings should be within 1 EV of one another. If they are not within 1 EV, the lights should be repositioned more symmetrically about the subject-to-camera line.

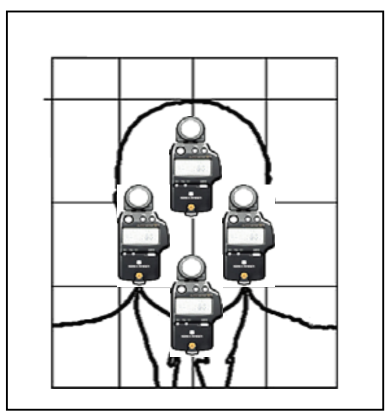


Figure 32 — Positions of incident light meter for exposure value measurement

EV is the value given to any combination of shutter speed and aperture (F-stop) that results in the same exposure. By definition, an EV value of 0 corresponds to a shutter speed of 1 second and an aperture of F1.0, for a film speed or equivalent image sensor sensitivity of ISO 100. EV is defined by the following equation:

$$\mathrm{EV} = Log_2\left(\frac{\mathrm{F}^2}{\mathrm{T}}\right) = 2Log_2(\mathrm{F}) - Log_2(\mathrm{T}),$$

where F is the F-stop setting and T is the exposure time. A change of 1 EV corresponds to a one F-stop aperture increase or decrease or a halving or doubling of the exposure time.

B.2.1.3 Example configurations for a photo studio or store

Described below are three examples of lighting and subject and camera positioning that are applicable to photographic studio businesses, as well as for some photofinishers that might offer identification photographs, in addition to their main business of material sales and film developing and printing. Example 1 is a single-light arrangement in which the placement of a panel of reflective material is used to provide more balanced lighting. Example 2 is a two-light arrangement with a lower reflective panel providing illumination to the region under a subject's chin. Example 3 is the same as Example 2, but with a third light behind the subject to eliminate shadows on the background material. Several recommendations for camera and subject positioning are also provided below.

B.2.1.3.1 Example 1: Proper lighting arrangement with a single light

In this arrangement, illustrated in Figure 33, a single light and multiple reflector panels are employed to illuminate the subject's face uniformly. The light, shown with a lamp reflector, should be placed approximately 35 degrees above the line between the camera and the subject and be directed toward the subject's face at a horizontal angle of less than 45 degrees from the line. A reflector panel should be placed on the subject's opposite side to prevent shadows on the face. As an option, an additional reflector may be placed below and in front of the subject's face to illuminate the area around the chin.

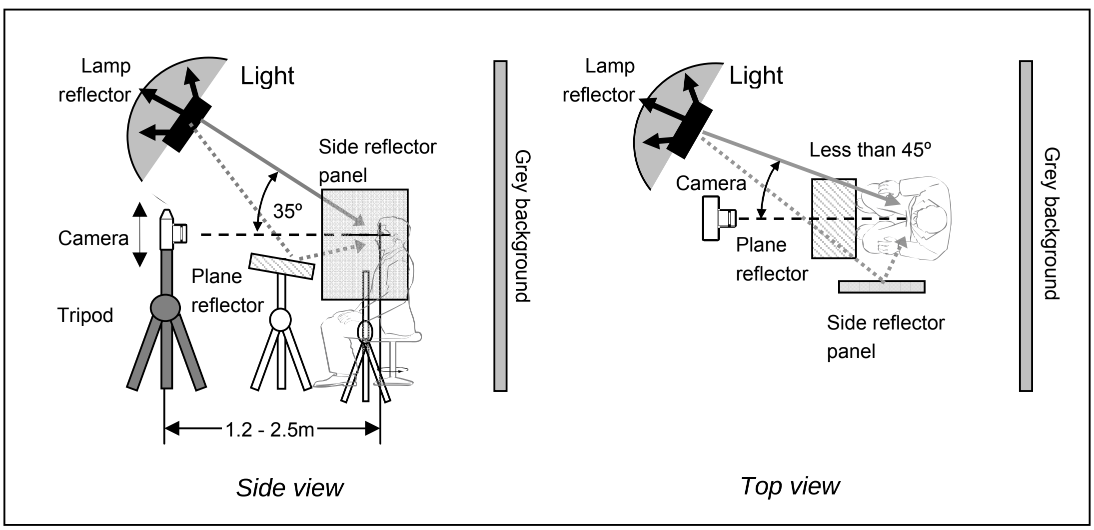


Figure 33 — Lighting arrangement for a photo studio with a single front light

B.2.1.3.2 Example 2: Proper lighting with dual lights

In the second example illustrated in Figure 34, two lights are employed. The lights, shown with lamp reflectors, should be placed approximately 35 degrees above the line between the camera lens and the subject. Both lights should be placed within 45 degrees of the line between the camera lens and the subject. Such an arrangement softens the edge of shadows and makes the lighting on the subject more even. The optional plane reflector in front of the subject supplies additional light around and below the subject's chin.

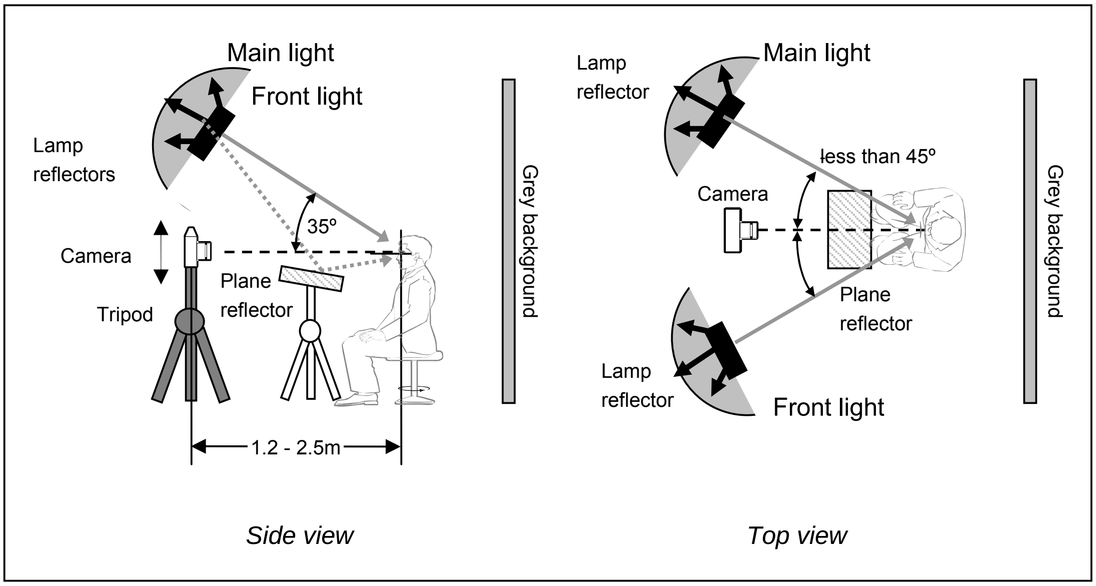


Figure 34 — Lighting arrangement for a photo studio with dual front lights

B.2.1.3.3 Example 3: Proper lighting with dual lights and background lighting

The use of a background light added to the arrangement shown previously in Example 2 should eliminate shadows visible on the background behind the face. As illustrated in Figure 35, the background light should be aimed at the background and be placed directly behind and below the subject.

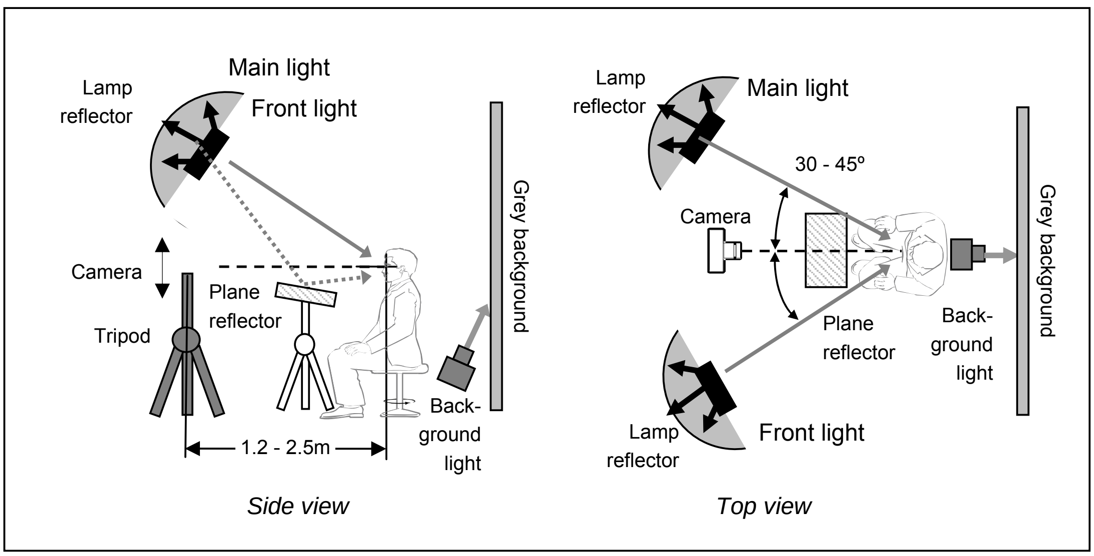


Figure 35 — Lighting arrangement for a photo studio with dual front lights and a background light

B.2.2 Recommendations for photo booths

A photo booth is typically a coin-operated, self-portrait photography unit, mostly used for taking ID pictures and equipped with such tools as a camera, lighting, stool, plain background, printing device and monitoring screen, and sometimes including an audio self-guidance application. Optimizing photographic technology enabled its space-saving size, which has contributed to its widespread use around the world. Following are some guidelines for the design and operation of such photo booths. Front, side, and top views of the arrangements described in the guidelines are provided. Also provided in this clause are suggestions for camera and subject positioning and a description of methods to provide feedback to the subject concerning his or her pose and expression.

B.2.2.1 Proper lighting

- Position multiple lights behind a diffuser panel and symmetrically above the camera. This will provide even lighting on the subject’s face and eliminate most glare and shadow problems. Place a background light low and midway between the background and the subject.
- The placement of the front lights 35 degrees above the line between the camera and the subject's head prevents direct reflection of the flash from a subject’s glasses.
- The inside walls should be white, except directly behind the subject. The white walls serve as reflectors and ensure that lighting on the face is uniform horizontally and vertically.
- The interior lights of the booth should be left on during operation. This will usually eliminate red-eye problems associated with photography in dim light.
- To eliminate unwanted shadows around the chin caused by lights above the subject, direct or indirect lighting from below and in front of the subject should be used.
- To ensure that the booth is free from the effects of external light, an opaque curtain should be employed.

B.2.2.2 Example configuration for a photo booth

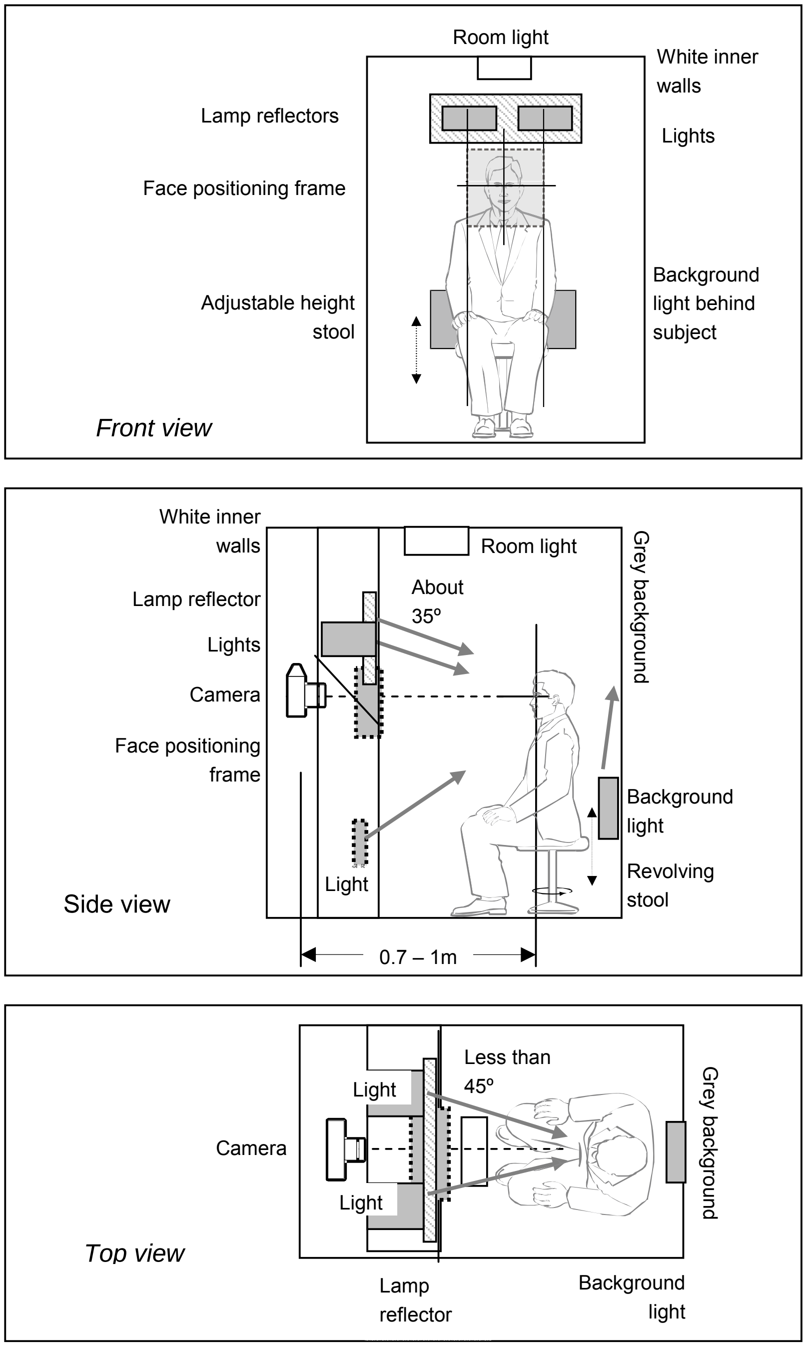


Figure 36 — Recommended placements of subject, camera, and lights in a photo booth

B.2.2.3 Camera-subject positioning

Proper positioning of the subject and control of the subject's pose can be improved through feedback provided to the subject via a mirror or a live-video monitor.

- A display device should be installed in the booth to provide a live image of the subject on the wall he or she faces. The device could be a one-way (half-silvered) mirror or a left-right reversed live-video monitor. The display should contain a frame which the subject can use to ensure that his/her entire head is fully visible, that his/her eyes are at the correct height, and that his/her face is cantered in the camera's field-of-view. Such a frame is illustrated in the following diagram.

- A height-adjustable chair or stool should be provided to allow the subject to face the camera and adjust his eyes to the proper height.

- Camera-to-subject distance is generally within 0,7-1,0 m.

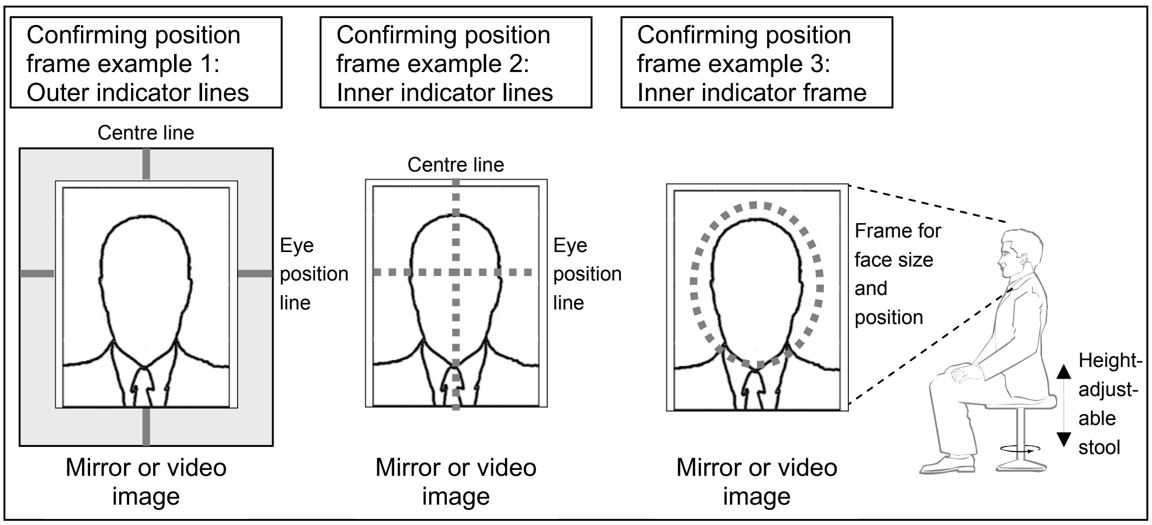


Figure 37 — Use of a display frame for head positioning

B.2.2.4 Adjustment of size, expression, etc. by monitor-GUI

- An image preview should be provided to allow a subject to recapture the image before it's printed or written to a storage medium, in case a subject might deem his/her pose or expression unacceptable. Illustrations of acceptable poses and expressions should be provided inside the booth.

- The size of the head in the image should be adjustable before printing or storage by allowing the subject to identify the positions of his/her crown and chin in a preview image. The system would then scale and crop the image accordingly. An illustration of such a preview image is provided in Figure 38.

- Alternatively, face detection software that automatically sizes and centres the head within the field-of-view can be used to ensure proper head positioning. Given that such software sometimes does not determine the face position correctly, a preview image should be provided with provision for manual override of the automatically determined position.

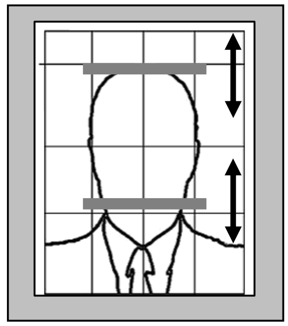


Figure 38 — Use of movable horizontal lines to set head size

B.2.3 Recommendations for a registration office environment

One of the major application areas of this standard is the use in the context of identity cards and especially Machine Readable Travel Documents (MRTDs). In this domain, images of applicants are either supplied by the applicant or taken in a “live enrolment scenario” in the registration office. For the first case, sub clauses B.2.1 and B.2.2 apply. In contrast, this sub clause focuses on the recommendations for live enrolment in a registration office environment, where lack of space often is a major concern. Nonetheless, image quality should be as close as possible to that achieved through the recommendations provided previously for the photo studio or photo booth, particularly for MRTDs.

B.2.3.1 Proper lighting

In the best practice arrangement for the Registration Office Environment illustrated in Figure 39 the subject and the background are illuminated by two diffuse light sources that are mounted in a console with a small footprint, so that it fits into a typical registration office environment. The console can be mounted on the floor or on the wall. Synchronized flash type illumination should be used for these light sources to enable high shutter speed and, thereby, avoid blur caused by a subject's motion. Direct lighting from the sun should be prevented in the office by proper means, e.g., by using curtains or roller blinds.

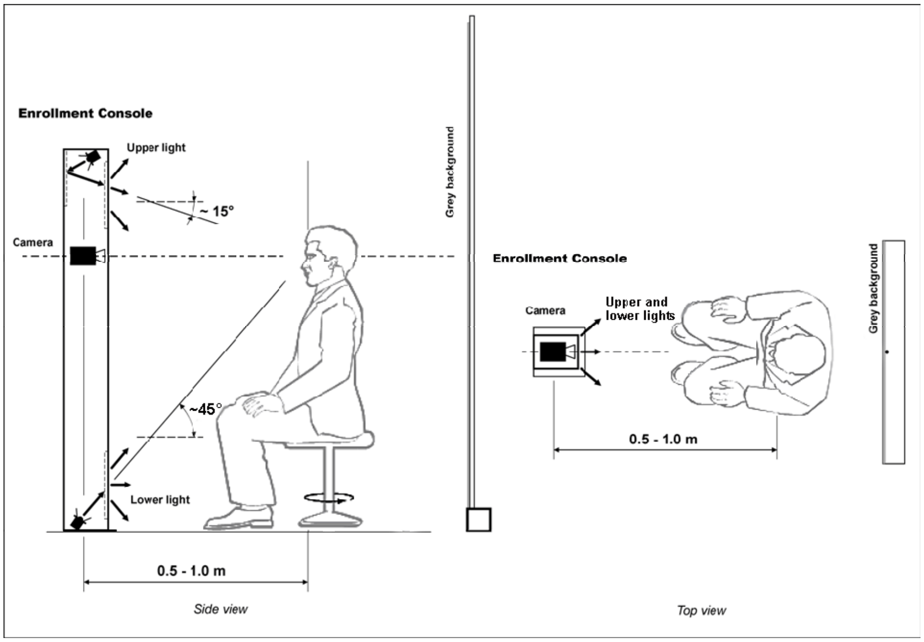


Figure 39 — Recommended placements of subject, camera, and lights in a registration office environment

B.2.3.2 Camera-subject positioning

- Proper positioning of the subject and control of the subject's pose can be improved through feedback via a second live-video monitor facing to the subject. An image preview should be provided to allow a subject to be recaptured before the image is submitted for further processing, in case a subject might deem his/her expression unacceptable.
- A revolving and height-adjustable chair or stool with an additional cushion for little children should be provided to allow the subject to face the camera and adjust his eyes to the proper height.
- Camera to subject distance is generally within 0,5 to 1,0 meter.

B.2.3.3 Operator support

- Proper positioning of the subject and control of the subject's pose can be improved through feedback to the operator via a live-video monitor showing the face of the subject during the acquisition process. An image preview should be provided to allow the subject to be recaptured before the image is submitted for further processing, in case the operator might deem the pose or expression unacceptable.
- To further improve the process and before the image is further processed, the operator should approve the quality of the image. This can be done with the support of face image quality assessment software, which enables checking the image automatically to ensure it fulfils the requirements of the registration authority or the requirements and best practices provided in this standard. Clause B.5 provides further information on such software and two examples of user interfaces.

B.3 Guidelines for printing

If used for the submission of images to a face recognition system, printing of the images (by either traditional photographic or newer digital techniques) will often be the step between acquisition and recognition that most limits image quality. Thus, it is important that the inherent capabilities of the printing process be used optimally to preserve the maximum amount of information in the image.

Digital printing systems can be broadly categorized into two types: (1) those that are able to control either the density or dot size of ink deposited at each ink dot; and (2) those that can produce ink dots only of a consistent size and density but that can change the frequency of occurrence of the dots. The first category can be termed continuous-tone printers, the second half-toning printers. If of sufficient resolution, either type may be used to produce prints for scanning and submission to automated face recognition.

B.3.1 Spatial and tonal resolution trade-offs

In any printing method, there is necessarily a trade-off between spatial and tonal resolution (or intensity gradation)—that is, a trade-off exists between the fineness of image details that can be preserved and the smoothness of transitions between pixel levels. Ideally, a process employed for printing face images should not limit either spatial details or tonal gradations to the extent that recognition system accuracy is diminished.

Many countries now expect to receive passport and visa photos that are approximately 50 millimetres square, with a typical distance between the eyes of about 12 millimetres, assuming that face height is somewhat greater than half of the picture height. To provide the 120 pixels between the eyes recommended in Informative Annex A.3, such a photo would require scanning at a spatial sampling frequency of about 12 pixels per millimetre (300 pixels per inch). A halftone printer would need to be capable of printing about 160 dots per millimetre to preserve information at that level of spatial detail with 256 levels of tonal resolution—a level of printing accuracy achievable today only at considerable cost. Clearly, some compromises in spatial or tonal resolution, or both, will be required.

B.3.2 Recommended printing quality

Because the range of available printing technologies is so great, a full set of recommendations covering all potential techniques cannot be provided in this annex. Moreover, even for any one printing technique, the choice of paper can have a substantial effect on image quality. Therefore, the approach taken in the following sub clause is not to provide recommendations that are specific to a particular printing technology, but rather to describe the visual appearance of the resulting prints and the minimum quality of digital images resulting from scanning of the prints.

B.3.2.1 Resolution and posterization

Any face printing process should produce a smooth image that is capable of accurately rendering fine facial details, such as wrinkles and moles, as small as one millimetre in diameter on the face. All flesh tones from both light- and dark-complexioned subjects should be printed accurately and no “hot spots” or shadow drop-out should be apparent. Smooth facial details should be rendered without posterization or contouring. Posterization occurs when otherwise smooth details in an image are rendered as an abrupt change in printed colour or density (i.e., as a visible contour).

B.3.2.2 Saturation

With the exception of glare or glints caused by small areas of possible specular (mirror-like) reflection, no portion of the printed image should be saturated in white or black. In other words, no portions of the background or the subject’s garments should be printed fully white and details should be apparent in dark shadow regions.

B.3.2.3 Moiré or visible dot patterns

Digitization of printed photos often introduces artefacts, such as moiré, and certain printing processes can exacerbate the generation of such artefacts. The printing process employed should not produce any noticeable moiré pattern when its prints are scanned with a document scanner at a sampling frequency of 12 pixels per millimetre (300 pixels per inch) or lower in each axis. If a printed photo has been produced through a periodic half-toning process, scanning the photo will almost invariably introduce moiré. Thus, those printers, such as inkjet and laser printers, which inherently employ half-toning to simulate continuous tones, should use non-periodic (or dithered) half-toning methods. Furthermore, the printing process should not produce dot patterns visible to the unaided eye.

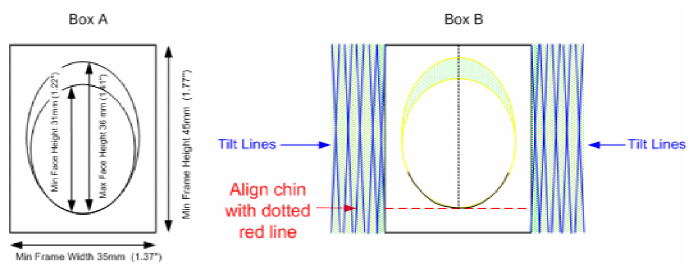
B.3.3 Use of a photo template

To facilitate evaluation of printed photographs, it is often useful to provide a transparent template to an acceptance agent or other individual charged with evaluating photo quality. The template would display the limits of head size and rotation (roll) and, when superimposed on the photo, could assist in the determination of whether a submitted photo is compliant with the requirements. An example of such a template, provided by

Citizenship and Immigration Canada, along with instructions for its use, is reproduced (at a reduced size) in the figure below.

Photo Template

This template can be used to measure head size, frame size, and degree of head tilt of the photo.



This template can be used to measure head size, frame size, and degree of head tilt of the photo.

Instructions

Head Measurements:

1. **Align the bottom of the ovals with the chin. Measurement of the head is from the chin to the crown[1]**
2. **If the head size is smaller than the 31mm oval (small oval) or larger than the 36 mm oval (large oval) the photo should be rejected. If crown is within the green shaded area, 31mm to 36mm, head size requirement is met for the photo.**

[1]Crown: The top of the head, or (if obscured by hair or head covering), where the top of the head or skull would be if it could be seen.

Frame Measurements:

1. **Align the bottom left side corner of the template frame line with the bottom left side corner of the frame of the photo.**
2. **If the frame size is smaller than in the defined standards of 35mm X 45mm, the photo should be rejected.**

Tilt Measurements:

1. **Align the bottom of the ovals with the chin. Tilt the photo so that the middle of the chin is in line with the middle of the nose, using the centre vertical black dotted line.**
2. **If the frame of the photo is tilted outside the blue tilt lines, the photo should be rejected. If the frame of the photo is within the green shaded area, the tilt requirement is met for the photo.**

March 2006 – Version 3

Figure 40 — Example photo template (provided by Citizenship and Immigration Canada)

B.4 Guidelines for scanning

The intent of these guidelines for scanning is to preserve, to the extent possible, all face-identifying information present in a printed photo when it is converted to a digital image. Preserving the information in a printed colour photo involves sampling it at a sufficient spatial frequency with adequate quantization in at least three spectral regions, usually in the red, green, and blue regions of the spectrum.

B.4.1 Sampling frequency and quantization levels

For a typical passport or visa photo a minimum spatial sampling frequency of about 12 pixels per millimetre (300 pixels per inch) is required to provide about 120 pixels between the eyes. The number of quantization levels should be at least 256 levels per colour, with three colours per pixel.

B.4.2 Spatial resolution

Spatial resolution is a measure of the ability to discern fine detail in an image and, although it's related, it is a metric distinct from sampling frequency. An often used measure of spatial resolution is the modulation transfer function (MTF). To ensure fine facial details are preserved, the scanner's MTF should be at least 20 % at a spatial frequency of 6 cycles per millimetre. Furthermore, the scanner's spatial resolution should be very similar in both axes and should not be enhanced through image sharpening or high-pass filtering algorithms.

B.4.3 Output colour space

Since red-green-blue (RGB) colour space and its derivatives are inherently device-dependent, the scanner's output should be converted to a well-defined, device-independent colour space such as *sRGB [19]*. Alternatively, an International Colour Consortium (ICC) standard-compliant colour profile for the scanner's native space can be embedded within the output image file.

B.4.4 Saturation

Saturation occurs when significant numbers of pixels have values that are at the limits of quantization, i.e., at the levels of 0 or 255, if quantization of eight bits per colour is employed. Acceptable scanned face images should not have a significant number of pixels in saturation.

B.4.5 Image compression

To reduce scanned image files to manageable sizes, lossy image compression with JPEG or JPEG2000 is typically used in document and film scanners. For some scanners, the manufacturer or software designer has incorporated a single or limited number of image quality-compression ratio trade-off points that a user may choose. The selection of overly high compression may substantially reduce image quality and, hence, diminish the accuracy of face recognition, as described in sub clause A.3.3 of Informative Annex A of this standard. Therefore, a user must be cognizant of the available image quality alternatives offered by the scanning system and the consequences of selecting certain quality-size trade-off points. If storage space or transmission time for the generated images is not overly limited, an alternative offering high image quality and, hence, large file size is an appropriate choice to ensure best possible recognition accuracy.

B.5 Face image quality assessment software

Several products, in the form of software development kits, which attempt to measure the compliance of submitted face images with various requirements described in the body of this standard, have become available commercially. Typically, these products provide multiple measures of face image quality, as well as a single, combined overall quality and can determine automatically whether submitted face images are likely to be of adequate quality for a particular application. Some of the metrics which have been incorporated into these products include face size, face centering, contrast, focus, background texture or uniformity, lighting uniformity, and head rotation (yaw and roll). Ideally, these metrics will correlate well with human perceptions of quality or with automated face matcher performance and will be thresholdable at user-defined settings. Face image quality assessment software may be a useful adjunct to human assessments of quality in photo studios, photo booths, or registration office environments. Examples of the graphical user interfaces (GUIs) for two such products are shown in the following figure.

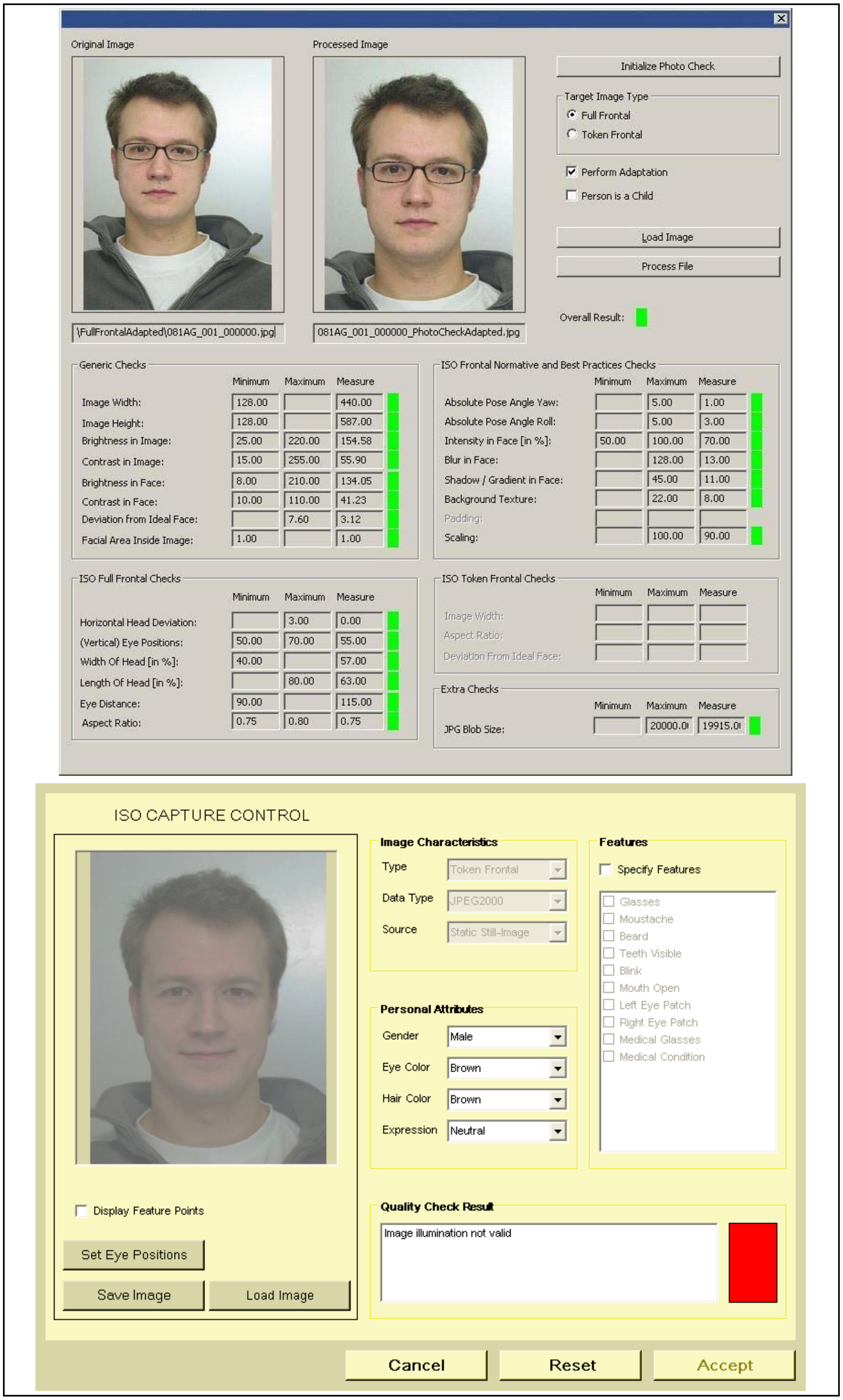


Figure 41 — Two sample GUIs of face image quality assessment software

B.6 Tables of the recommendations

Most of the recommendations for scene setting, photographing, after photographing and photographic quality are summarized in the following tables. In these tables, the Contents column contains references to the relevant sections of this standard. The **PS** and **PB numbers** relate to the similarly labelled photo studio and photo booth examples provided in Clause B.7, Photographic Examples.

B.6.1 Scene setting

Table 41 — Scene setting Recommendations

Category	Contents	Photo studio	Photo booth	Registration office
Lighting	Number of lights (ref. 7.2.8 Hot spots)	Use 2 or more lights, if possible. When a single light is used, employ reflectors to redirect the light. Background lighting is recommended to reduce shadows. **(PS1)**	Use 2 or more lights, if possible. When single light is used, white walls that can serve as reflectors are recommended. Background lighting is recommended to reduce shadows. **(PB1)**	Use 2 or more lights. If possible, background lighting is recommended to reduce shadows on background.
	Lighting methods (ref. 7.2.8 Hot spots)	Provide diffuse illumination by using lamp reflectors (ex. bounce umbrella), lamp diffusers(ex. soft box strobe), reflector panels, etc. **(PS2)(PS3)**	Provide diffuse light using lamp reflectors or a white translucent plastic panel. White walls can also serve as reflectors. **(PB2)(PB3)**	Provide diffuse light using white translucent plastic panel for down and up light.
	Colour temperature	4 500 — 6 500K is recommended.	4 500 – 6 500K is recommended. Hold colour balance using grey or white walls. Eliminate stray or outdoor light with a curtain or door.	4 500K – 6 500K is recommended. Hold colour balance using grey or white background. Minimize outdoor light by using appropriate means (e.g. curtains).
	Guide Number of strobe	Recommended Guide Number is over 7 for external flash, 14 for built-in flash. (photographic distance of 1,2 – 2,5 m) The built-in flash in compact cameras is not recommended because the Guide Number is typically only about 10, so it is insufficient, and it produces bad lighting (e.g., shadows on the background and red-eye)	Guide Number should be over 7. (photographic distance of 0,7 – 1,0 m)	Guide Number should be over 10 (photographic distance of 0,5 – 1,0 m). Use two balanced strobes.
	Lighting Uniformity (ref. 7.2.7 Subject and scene lighting)	Lighting should be evenly distributed. The difference of four exposure values on the left and right sides of a face, chin and forehead should be less than 1 EV. **(PS4)**	Lighting should be evenly distributed. The difference of four exposure values on the left and right sides of a face, chin and forehead should be less than 1 EV. **(PB4)**	Lighting should be well balanced. The difference of four exposure values on the left and right sides of the face, chin and forehead should be less than 1 EV.
	Light arrangement	30 – 45 degrees of subject-lens axis. 35 degree elevation angle for subject-light line. **(PS5)**	Symmetrical position. 30 – 45 degrees to subject-lens axis. 35 degree elevation angle for subject-light line. **(PB5)**	Above and under the subject-lens axis. Typically 15° degrees from above and 45° degrees from below.

Category	Contents	Photo studio	Photo booth	Registration office
Subject	Pose angle less than +/- 5 degrees) (ref. 7.2.2 Pose)	Provide simple guidance to the subject concerning the proper pose. Position camera horizontally on the centre line of the face, vertically between the eyes and the tip of the nose. Camera height adjustment may be done with the tripod. **(PS6)**	Guide the proper position of the face for the subject. Position camera horizontally on the centre line of the face, vertically between the eyes and the tip of the nose. The height adjustment can be done with a revolving stool for the subject. The subject can also check the proper position by a mirror in front of him/her or by a preview image provided before or after image capture. Expression guidance is recommended to place on elsewhere of photo booth. **(PB6)**	Provide simple guidance to the subject concerning the proper pose. Camera position meets horizontally right at the centre line of the face, vertically between eyes and the tip of the nose. The height adjustment can be done with a revolving stool for the subject and an additional cushion for small children. The operator as well as the subject can check proper position on (a) preview monitor(s) if provided.
	Shoulder positioning (square on to the camera) (ref. 7.2.5 Shoulders)	Give the subject advice. **(PS7)**	The subject can also check the proper position by the mirror in the front of him/her or the image shown before/after the shoot. Indicate the suggested position of shoulders on the inner wall in front of the subject. The photograph can be retaken in case it is unacceptable. **(PB7)**	Give the subject advice. The operator as well as the subject can also check the proper position of him/her or the image shown before/after the shoot in case a preview is provided. The photograph can be retaken in case it's unacceptable.
	Face size adjustment	In case of no size adjustment after photographing, face size or face position (crown and chin) should be confirmed before photographing on the screen of camera. The adjustment can be made by, if needed, adjustment of (1) distance between a subject and camera or zoom magnification for face size and (2) height and position of camera for face position.	In case of no size adjustment after photographing, face size or face position (crown and chin) should be confirmed by the subject before photographing on the monitor or a mirror which is equipped. The adjustment can be made by, if needed, adjustment of (1) zoom magnification for face size and (2) height of a stool or a chair for face position.	Face size or face position (crown and chin) should be confirmed on the operator preview and especially on the output of the quality assessment software presented to the operator. If needed, the adjustment can be made by the distance of the subject and the camera or the zoom magnification handled by the operator for face size. For the vertical position a necessary adjustment is done by the revolving stool.
	Recommended Expression (Neutral, non-smiling, with both eyes open and mouth closed) (ref. 7.2.3, A.2.4 Expression)	Give the subject some advice. **(PS8)**	The subject can also check the proper expression by the mirror in the front of him/her or the video image shown before/after photographing. The photograph can be retaken in case it is unacceptable. **(PB8)**	Give the subject some advice. The operator as well as the subject can check proper expression on the preview monitor if provided. The photograph can be retaken in case it's unacceptable.
	Eye glasses (Acceptable when: Normal use, Permanently tinted for medical reason, Frames do not obscure the eyes) (ref. 7.2.9 Eye glasses)	Give the subject some advice. Adjust the angle between the light and the subject to reduce glare. **(PS9)**	The subject can also check the glare by the mirror in the front of him/her or the image shown before/after the shoot. The photograph can be retaken in case it is unacceptable. **(PB9)**	Give the subject some advice. The operator as well as the subject can check the glare the preview monitor if provided. The photograph can be retaken in case it's unacceptable.
Background	Background colour (ref. A.2.9.4 Background examples)	Grey, light blue, white, off-white	Grey, white, off-white	Grey, light blue, white, off-white

Category	Contents	Photo studio	Photo booth	Registration office
	Background uniformity (ref. A.2.9.3 Background uniformity)	Plain. A uniform or a single colour pattern with gradual change caused by background light	Plain. A uniform or a single colour pattern with gradual change caused by background light	Plain

B.6.2 Photographing

Table 42 — Photographing recommendations

Category	Contents	Photo studio	Photo booth	Registration office
Camera set-up	Shutter speed	Shutter speed (1/60-1/250) should be fast enough to prevent motion blur, unless electronic flash is the predominant source of illumination. Camera should be fixed with a device such as a tripod or rigid stand. **(PS10)**	Shutter speed (1/60-1/250) should be fast enough to prevent motion blur, unless electronic flash is the predominant source of illumination.	Shutter speed (1/60-1/250) should be fast enough to prevent motion blur, unless electronic flash is the predominant source of illumination.
	White balance (ref. A.2.14 Colour calibration)	Manual correction. Set up with the colour of grey in order to check the proper white balance. **(PS11)**	Manual correction at the time of Photo Booth installation or maintenance. Set up with the colour of grey in order to check for proper white balance.	Automatic correction before shooting.
	Over or under exposure (ref. 7.3.2 No over or under exposure)	Check with an exposure meter. Use 18% grey test chart for the proper adjustment of exposure. **(PS12)**	Correct exposure while setting system. Use an 18% grey test chart for the proper adjustment of exposure.	Correct exposure while setting up system. Use an 18% grey test chart for the proper adjustment of exposure.
	Camera-to-subject distance (ref. A.2.15 Radial distortion of the camera lens)	1,2 – 2,5 m in a typical photo studio **(PS13)**	0,7 – 1,0 m in a typical photo booth **(PB10)**	0,5 – 1,0 m in a typical registration office scenario
	Focal length of camera lens (35 mm format equivalent) (ref. A.2.15 Radial distortion of the camera lens)	A normal to a medium telephoto lens (50 – 130 mm) (Shooting distance of 1,2 – 2,5 m)	A normal to medium telephoto lens (40 – 100 mm) (Shooting distance of 0,7 – 1,0 m)	A light wide angle to medium telephoto lens (36 mm – 100 mm) (Shooting distance of 0,5 – 1,0 m)
	Resolution (ref. 7.3.3 Focus and depth of field)	Higher than 2 pixels per mm (Can be checked by test shooting a ruler)	Higher than 2 pixels per mm (Can be checked by test shooting a ruler)	Higher than 2 pixels per mm (Can be checked by test shooting a ruler)
	Face size adjustment (ref. 8.3 Photographic requirements for the Full Frontal Face Image Type)	Face position marks, indicated by feature points or bars at the positions of the subject's crown and chin bottom, should be visible in the camera's viewfinder screen. This allows the photographer to ensure that distance and zoom are correct before shooting.	A mirror or a live-video monitor that displays face position marks, indicated by feature points or bars at the positions of the crown and chin bottom, could be used. This will enable a subject to fit his face within the image properly and to make any needed adjustments, such as zoom lens magnification, before shooting.	A live-video monitor that displays face position marks, indicated by feature points or bars at the positions of the crown and chin bottom, could be used. This will enable the operator to make any needed adjustments, such as zoom lens magnification, before shooting.

B.6.3 After photographing

Table 43 — Printing and scanning recommendations

Category	Contents	Photo studio	Photo booth
Printing	Colour and density	Teeth and whites of eyes shall be clearly light or white and dark hair shall be clearly dark.	Teeth and whites of eyes shall be clearly light or white and dark hair shall be clearly dark.
	Resolution	Photographs should have a life-like quality. Resolution shall be sufficient to resolve facial details less than 1 millimetre in diameter.	Photographs should have a life-like quality. Resolution shall be sufficient to resolve facial details less than 1 millimetre in diameter.
	Size adjustment of face	Digital: Use application software to specify the position of the crown and chin. Analogue: Adjust zoom magnification mode while printing.	Digital: The position of crown and chin can be specified on the monitor. Analogue: Adjust zoom magnification mode before shooting.
Scanning	Colour and density (ref. 7.3.5 Colour or greyscale enhancement, 7.4.2.1 colour space)	Teeth and whites of eyes shall be clearly light or white and dark hair shall be clearly dark. 24 bit-RGB. Use a device-independent colour space or embed the device's colour profile. Retouching should not be done.	
	Spatial Sampling (ref. A.3.1.1 Photo Resolution)	About 12 pixels per millimetre (300 ppi).	
	Cropping (ref. 7.4.1.1 Pixel aspect ratio)	Pixel aspect ratio of 1:1 should be maintained.	
	Size adjustment of face image (ref. 8.3 Photographic requirements for the Full Frontal Face Image Type, A.3.2 Best practice for use of Full Frontal Images on travel document)	When necessary, use only bi-linear or bi-cubic methods or other superior resampling algorithms to resize or rotate the face image. Resampling should only be performed to decrease the image's pixel dimensions, not to increase them.	
Category	**Contents**	**Registration office**	
Printing and Scanning		Not Applicable, as live enrolment is done to avoid a breech in the digital production chain. Document printing on the ID document (e.g. passports, ID cards, driver's licenses, etc.) is driven by security requirements on the final document and the technology suited for fulfilling these requirements (e.g. laser engraving, etc.)	
Quality Assessment	Geometric requirements	The quality assessment software should check if the image and the head shown in the image (after cropping) meet the geometric requirements imposed by this standard and the requirements imposed by the registration authority. Inter-eye distance and positioning of the head are prominent examples that should be checked automatically.	
	Constraints on the subject	The quality assessment software should check if the image meets the requirements imposed on the subject (e.g., pose) by this standard and the requirements imposed by the registration authority.	
	Constraints on the scene	The quality assessment software should check if the image meets the requirements regarding the scene composition (e.g., background colour) imposed by this standard and the requirements imposed by the registration authority.	
	Constraints on the photographic requirements	The quality assessment software should check if the image meets the requirements regarding the photographic requirements (e.g., absence of blur) imposed by this standard and the requirements imposed by the registration authority.	

B.6.4 Photographic quality

Table 44 — Photographic quality recommendations

Category	Contents	Photo studio	Photo booth
Photographic quality	Evenness of illumination (ref. 7.2.8 Hot spot)	a) The difference of four Exposure Values on the left and right sides of a face, the chin, and forehead should be less than 1EV. **(PS4)** b) Light source should be diffused by using reflectors such as bounce umbrellas or reflective panels. Direct light should be dispersed in multiple directions. **(PS2)** c) Calibrate the correct exposure using an exposure meter. Do not allow over exposure.	a) The difference of four Exposure Values on the left and right sides of a face, the chin, and forehead should be less than 1EV. **(PB4)** b) Light source should be diffused by a diffuser panel or white walls serving as reflectors. Direct light from the front should be dispersed in multiple directions. **(PB2)** c) Do not allow over exposure.
	Shadows over the face (ref. 7.2.12 Lighting artefacts)	a) It is preferable to direct the main light toward the subject's face at a horizontal angle of less than 45 degrees. **(PS5)** b) If there's any shadow, place a reflector panel in front of the shadow so that the light is bounced off its reflective surface to lighten the shadow and soften contrast. c) The difference between Exposure values on left and right sides of a face, or chin and forehead should be less than 1EV. **(PS1) (PS4)**	a) It is preferable to direct the main light toward the subject's face at a horizontal angle of less than 45 degrees from the camera to subject axis. Should use more than 2 lights. b) The difference of Exposure values between left and right sides of a face, or chin and forehead should be less than 1EV.
	Background shadows (ref. A.2.9.2 Background shadows)	Use a background light to eliminate shadows visible on the background behind the subject.	Use a background light to eliminate shadows visible on the background behind the subject. White walls will reflect some of the light and lighten the shadows.
	Glare on glasses (ref. 7.2.9 Eye glasses)	Direct the light toward the subject at a vertical angle of more than 35 degrees. This will reduce the glare. **(PS9)**	Direct the light toward the subject at a vertical angle of more than 35 degrees. This will reduce the glare. The subject can check the proper position by the mirror in the front of him/her or a video image shown before or after photography. The photograph can be retaken in case it's considered unacceptable. **(PB9)**

B.7 Experimental data

B.7.1 Experimental results of face recognition in a photo studio and photo booth

Face recognition experiments were performed under several ideal and actual lighting conditions, including down lighting at airports that cast strong shadows on faces. The experiments were performed in Japan and supported by the Japanese Ministry of Foreign Affairs, with the participation of several face recognition vendors. The experiments were done under the practical requirements specified in this standard, such as object conditions (pose, expression, glasses, etc.), photographic equipment (camera, lighting, background, etc.) and other aspects in four different lighting conditions: best practice, bad airport lighting, improved lighting at airport, and photo booth.

The total number of the subjects photographed was approximately 160. The images were taken under the conditions listed in Table 45, with the passport photo width-to-height ratio of 1:1.29 and pixel dimensions of

420 × 540, which is the size typically used by the Japanese Ministry of Foreign Affairs. The images used in these tests were compressed to an average size of 80k bytes using JPEG compression. For each lighting condition tested, all images were compared to the best practice images in one-to-many (i.e., Identification) testing, using several face recognition products. The performance metric measured was the Cumulative Match Rate (CMR) at Rank 1.

Table 45 — Lighting arrangements of photographing conditions

Lighting conditions	Lighting	Photographing equipment	Shooting distance
Best practice	Front lighting from three directions (Diffused flash light sources) with a reflector panel placed under a subject's chin	Single-lens reflex (SLR) digital still camera f = 60 mm (f = 35-70 mm zoom lens)	2,4 m
Bad airport lighting (dark lighting and shadowed)	Overhead down lighting, ordinary light source, no reflector panels	SLR digital still camera f = 60 mm (f =35-70 mm zoom lens)	1,7 m
Improved lighting at airport	Single-point front lighting and overhead down lighting, ordinary light source, no reflector panels	SLR digital still camera f = 60 mm (f = 35-70 mm zoom lens)	1,7 m
Photo booth	Multiple lighting sources, from at least three directions, including indirect light source (diffused flash light source), and one background lighting	Fixed-focal-length digital still camera (f = 40 mm)	0,7 m

Following are the results from the experiment, which are also summarized in Table 46:

— In the comparison between best practice lighting and airport lighting, the CMR for typically bad airport lighting was fairly low: 40 – 80 %.

— To improve face recognition accuracy, lighting arrangements to reduce excessive shading on the face (i.e., improved lighting at airport) worked effectively.

— No significant difference in the CMR was found between the results for best practice and photo booth.

Table 46 — Comparison of results of face recognition accuracy for various lighting conditions

Lighting conditions enrolled image vs. query image	Cumulative Match Rate (%)
Best practice vs. bad airport lighting (insufficient intensity and shadowing)	40 – 80
Best practice vs. improved lighting at airport	Approx. 100
Best practice vs. photo booth	100

NOTE Cumulative Match Rate (CMR) refers to the Rank 1 match rate. The statistical uncertainty in the measured face recognition accuracy rates was approximately 3 %. Uncertainty refers to B.1.1 Rule of 3 in ISO/IEC 19795-1.

B.8 Photographic examples

To illustrate and explain the recommendations provided in the tables of Clause B.6, the following photographic examples were taken under various conditions in a photo studio and a photo booth. Preceding the examples are brief descriptions of the recommended conditions for a photo studio or photo booth and the specific camera settings chosen.

B.8.1 Photographic examples at a photo studio

a) Recommended photographic conditions at a photo studio

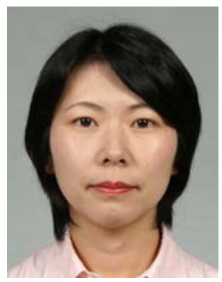
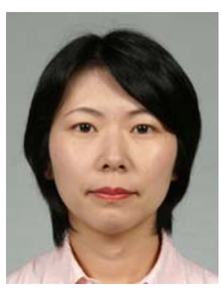
Table 47 — Recommended photographic conditions at a photo studio

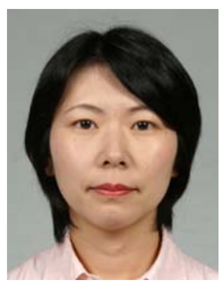
Category	Lighting	Subject	Background	Camera conditions
Photo-studio (See 2.1 Recommen-dations for a Photo Studio or Store)	Two lights, each diffused with a lamp reflector, single background light. Lights symmetrically arranged at 45° from camera-to-subject line. The angle of elevation is 35°. ΔEV = 0,2. (Horizontally 0,2 EV, vertically 0,2 EV), Front reflector, Colour temperature is 5 200K.	The photographer should provide instructions to the subject. For positioning adjustment, use a guide frame on the preview screen of the camera. Height adjustment by tripod or stool. Camera-to-subject distance is 1., m.	Grey	Manual, F11, 1/125 sec., f = 75 mm, measurement with exposure meter and colour temperature meter (f: equivalent focal length for 35 mm film camera)

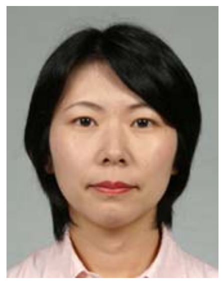
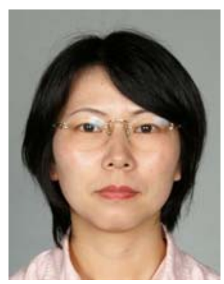
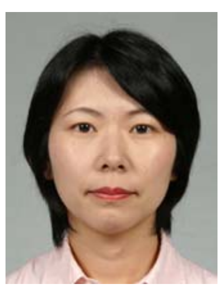
b) Photographic examples at a photo studio

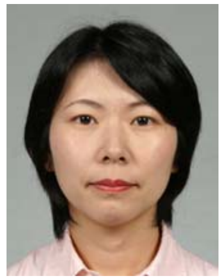
[NR]: Not recommended – image obtained under not recommended conditions.

Category	Item	Photographic examples		
Lighting	Number of lights **(PS1)**			
		Two lights and background light	Single light and side reflector panel	Single light (without side reflector panel) [NR]
	Diffuse illumination **(PS2)**			
		Lamp reflector	Without lamp reflector [NR]	Camera strobe light [NR]

Category	Item	Photographic examples		
	Reflector panel **(PS3)**			
		Front reflector panel, two lights	Without reflector panel, two lights	
	Lighting uniformity (Exposure value difference between right and left) **(PS4)**			
		ΔEV=0,2	ΔEV=0,5	ΔEV=0,8
	Light arrangements **(PS5)**			
		Symmetrically 45°	left 75°, right 45° [NR]	
				
		Angle of elevation: 35°	Angle of elevation: 70° [NR]	

Category	Item	Photographic examples		
				
		Angle of elevation: 35°	Angle of elevation: 0°, glare on glasses [NR]	
Subject	Pose angle **(PS6)**			
		Camera height adjustment	Without camera height adjustment [NR]	
				
		With instruction	Tilting the bridge of the nose [NR]	Incorrect face direction (gaze to the right) [NR]
	Shoulder positioning **(PS7)**			
		Square on to the camera	Not square on to the camera [NR]	

Category	Item	Photographic examples		
	Expression **(PS8)**			
		With instruction	Without instruction (wrong eye direction) [NR]	
				
		With instruction	Without instruction (smile) [NR]	
	Glare on glasses **(PS9)**			
		Tilting her head forward	Without tilting her head forward (Note glare on glasses) [NR]	
Camera conditions	Shutter speed **(PS10)**			
		1/125 sec.	1/300 sec. (poor synchronization) [NR]	

Category	Item	Photographic examples		
	White Balance **(PS11)**			
		Auto White Balance should not be used because it can be affected by the colour of the subject's clothing.		
		Manual setting	Manual setting (red colour wear)	Auto [NR]
				
		Correct adjustment to 5 200K(standard)	Wrong adjustment to camera setting at relatively lower colour temperature [NR]	Wrong adjustment to camera setting at relatively higher colour temperature [NR]
	Exposure **(PS12)**			
		Proper exposure	1 EV Over [NR]	1 EV Under [NR]
	Camera-to-subject distance **(PS13)**			
		1,2 m (f = 50 mm)	1,5 m (f = 75 mm)	2,5 m (f = 130 mm)

B.8.2 Photographic examples at a photo booth

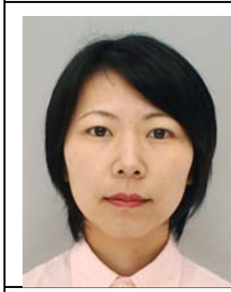
a) Recommended photographic conditions at a photo booth

Table 48 — Recommended photographic conditions at a photo booth

Category	Lighting	Subject	Background	Camera conditions
Photo Booth (B.2.2 Recom-mendations for Photo Booth)	Four diffused lights (single background light) Symmetrical arrangement at angle of about 25°, angle of elevation of 35°. ΔEV=0,4EV (horizontally 0,0EV, vertically 0,4EV), White walls as front and side reflectors. Colour temperature of 5 200K.	Audio self-guidance, half mirror with a frame for face position adjustment and confirming the face image after taking photos. Height adjustment by stool. Camera-to-subject distance is 0,7 m.	Grey	F5,6, 1/125 sec, f = 40 mm, measuring with exposure meter and colour temperature meter at system installation. (f: equivalent focal length for 35 mm film)

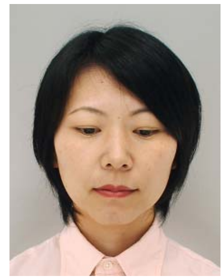
b) Photographic examples at a photo booth

[NR]: Not recommended – image obtained under not recommended conditions.

Category	Item	Photographic examples		
Lighting	Number of lights **(PB1)**			
		Three lights	Single light off (without light for chin) [NR]	
				
		With background light	Without background light [NR]	

Category	Item	Photographic examples		
	Diffuse illumination **(PB2)**			
		With diffuse illumination	Without diffused illumination (shadows) [NR]	
	Reflector panel **(PB3)**			
		With white walls reflector panels	Dark walls without reflector panels [NR]	
	Light uniformity **(PB4)**			
		Symmetrical light uniformity	Without symmetrical light uniformity [NR]	
	Light arrangements **(PB5)**			
		Angle of elevation 35°	Angle of elevation 70°[NR]	

Category	Item	Photographic examples		
				
		Angle of elevation 35°	Angle of elevation 0°(glare on glasses) [NR]	
Subject	Pose angle **(PB6)**			
		Instruction by face position frame	Without face position frame (looking downward) [NR]	Without face position frame (left direction) [NR]
				
		Height adjustment by stool	Without height adjustment (upper direction) [NR]	
	Shoulder positioning **(PB7)**			
		Square on to the camera with the instruction	Without instruction (right direction) [NR]	

Category	Item	Photographic examples		
	Confirmation of expression **(PB8)**			
		Re-photographing	Without re-photographing (wrong eye direction) [NR]	Without re-photographing (closed eyes) [NR]
	Glare on glasses **(PB9)**			
		Decreasing the glare by tilting his head forward	Glare on glasses [NR]	
	Camera-to-subject distance **(PB10)**			
		0,7 m	0,5 m [NR]	

Annex C
(informative)

Experimental studies

C.1 Experimental study on the enrolment of full frontal images for travel documents Specifications and data used for the analysis

This Informative Annex describes a study, the results from which provide justification for the tolerances regarding inter-eye distance, relative horizontal position of the face, relative vertical position of the face, head to image width ratio and head to image height ratio.

C.1.1 Specifications and data used for the analysis

The parameters and tolerances used in this study were either

1. The strict tolerances as demanded by ISO/IEC 19794-5:2005 standard and ICAO (International Civil Aviation Organization) recommendations [8].
2. The relaxed tolerances as suggested by ICAO [9] for the real-world application of passport image enrolment.

The data used for this study was derived from a large scale sampling of e-passport photographs. The data were contributed by the governments of four high volume e-passport issuing States. All images used in this study were already accepted for the issuance of an e-passport in the respective countries.

The focus of the analysis performed was largely on whether typical Passport photos meet the key specifications set out in ISO/IEC 19794-5; in particular those with respect to

1. pose,
2. number of pixels between the eye centres,
3. relative horizontal position of the face,
4. relative vertical position of the face,
5. head to image width ratio and
6. head to image height ratio.

Table 49 — Real-world image datasets used for the analysis

Dataset Country	# of images	Pixel size (width x height)	Format
0	1 000	413 x 531	JPEG
A	1 988	384 x 480	JPEG
B	1 911	449 x 599	JPEG
C	2 229	416 x 536	JPEG

The data derived from these passport images are subsequently compared with each other and with the tolerances specified by this part of ISO/IEC 19794.

C.1.2 Experimental results

C.1.2.1 Inter-eye distance

This part of ISO/IEC 19794 specifies a minimum of 180 pixels of the width of the head for full frontal images. It corresponds to roughly 90 pixels between eye centres (ref clause8.4.1).

Figure 42 shows the distribution of the inter-eye centres for the four sample data sets.

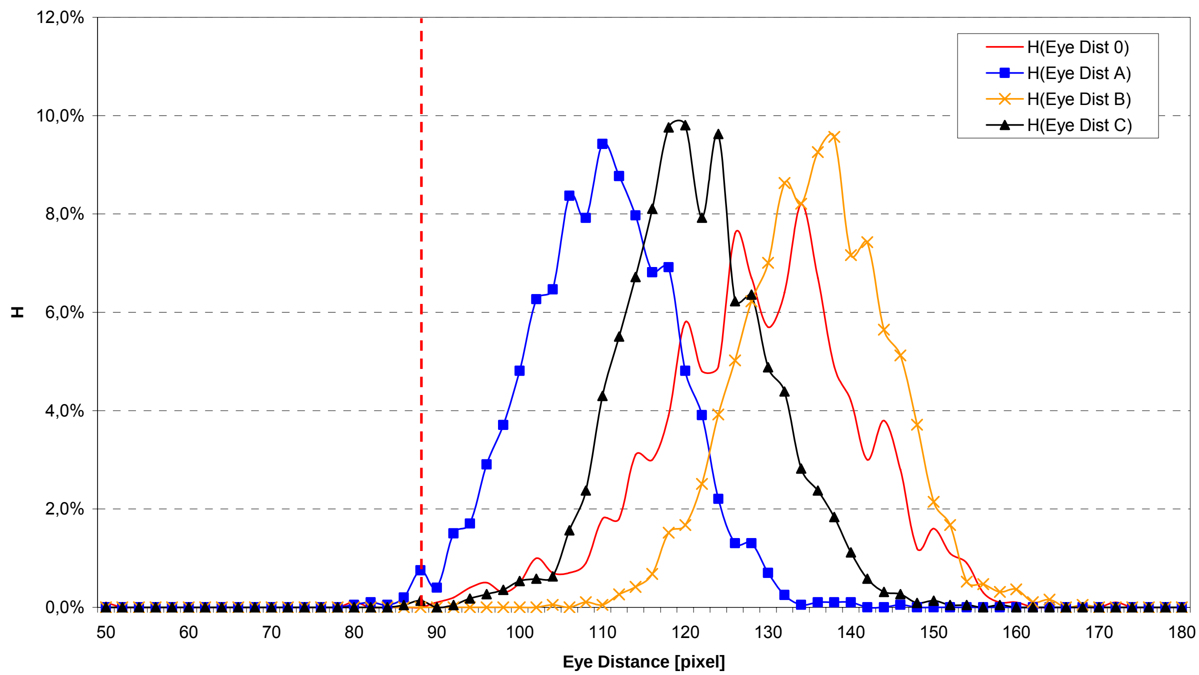


Figure 42 — Normalized distribution of the eye distance in the e-passport images of four Issuing States; "the red dot line represents 90 pixels roughly corresponding to the limit of full frontal images

The requirement of the width of head was met in almost all of the cases. The average distance between the eyes was found to be at 123,4 pixels.

C.1.2.2 Relative horizontal position of the face

This part of ISO/IEC 19794 specifies that the X coordinate M_x of the centre of the face M shall be between 45% and 55% of the image width (see clause 8.3.2). Figure 43 shows the distribution of the horizontal position of the face (M_x/A) for the four sample data sets.

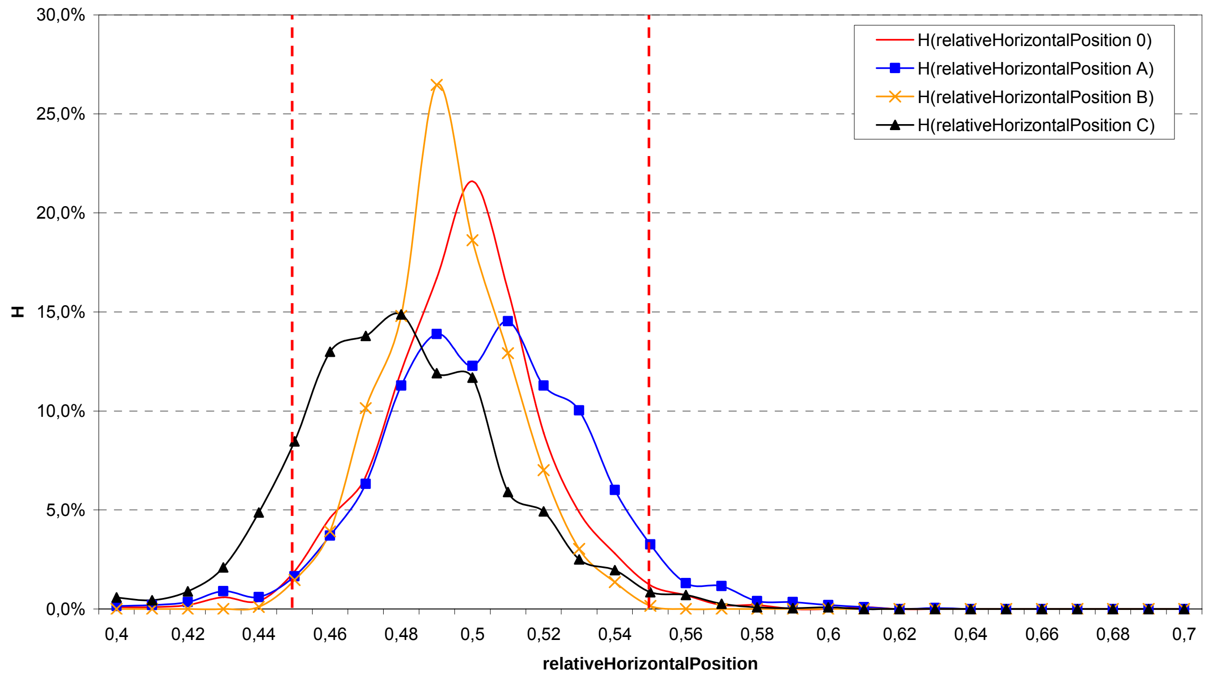


Figure 43 — Normalized distribution of the relative horizontal head position in the e-passport images of four Issuing States. The limits as specified by this part of ISO/IEC 19794 are depicted

The average horizontal head position of the 7 128 images is at approximately 49% of the image width. The specifications of this part of ISO/IEC 19794 have been met by 95,4% of all passport photos of this study.

C.1.2.3 Relative vertical position of the face

This part of ISO/IEC 19794 specifies that the Y coordinate M_y of the centre of the face M shall be between 30% and 50% of the image height *B* (ref. clause 8.3.3) with less strict requirements for children under the age of 11. Figure 44 shows the distribution of $1-M_y/B$ of the four sample data sets.

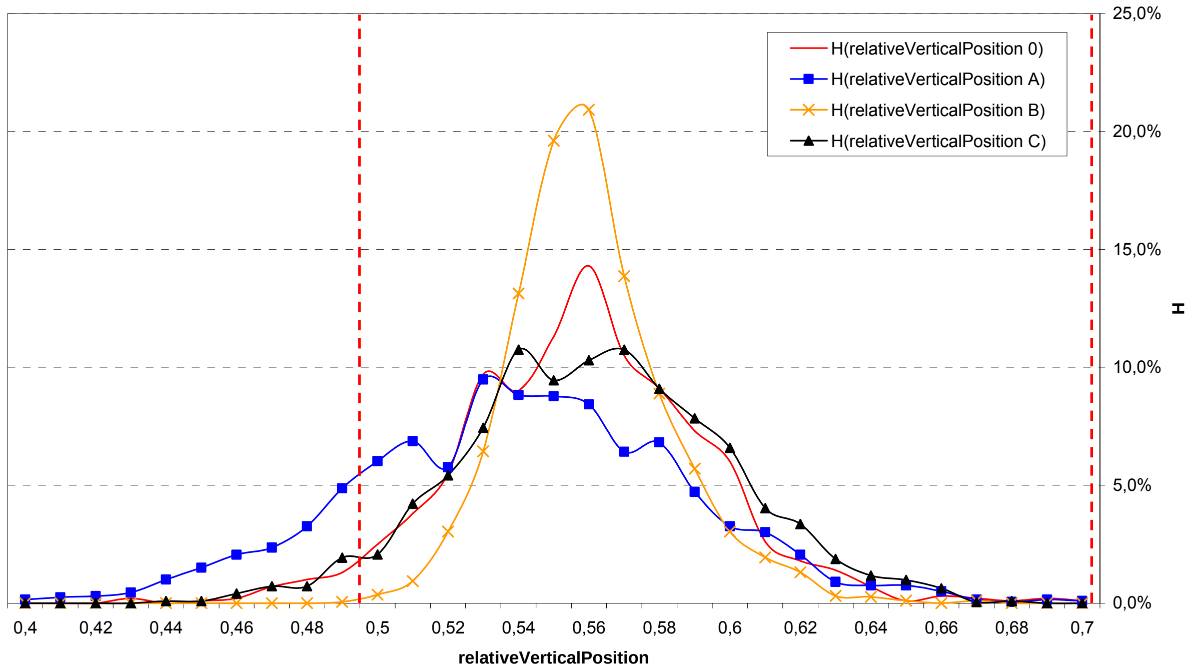


Figure 44 — Distribution of the relative vertical head position $1-M_y/B$ in the e-passport images of four Issuing States; the limits specified are depicted

The average vertical head position (vertical eye position, i.e. the position of a horizontal line through the centres of the eyes) of the 7 128 images $1-M_y/B$ is at approximately 0,56 or 56% of the image height. This corresponds to a Y coordinate of M (M_y) of 44% of the image height.

C.1.2.4 Width of the Head relative to the Image Width

In order to assure that the entire face is visible in the image the head width CC shall be between 50% and 75% of the image width (see clause 8.3.4). Figure 45 shows the distribution of the head width to image width ratio for the four sample data sets.

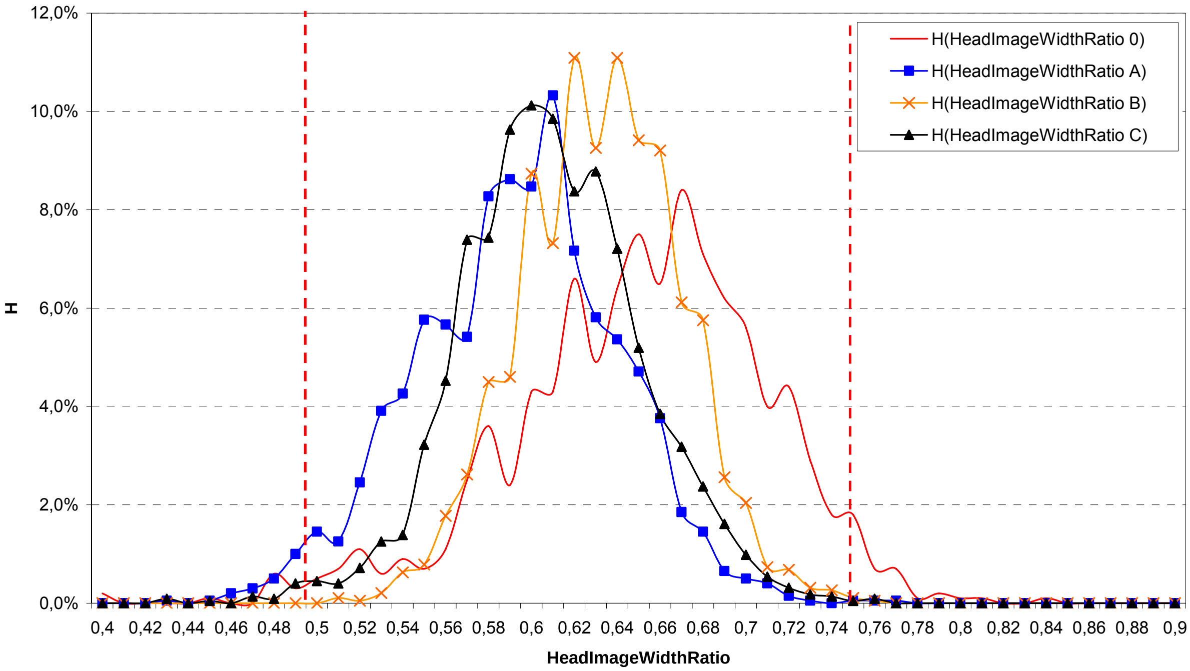


Figure 45 — Distribution of the head to image width ratio in the e-passport images of four Issuing States; the limits as specified in this part of ISO/IEC 19794 are depicted

The average head to image width ratio of the 7 128 images is found at 0,62. Most of the images of all four distributions meet the requirements of this part of ISO/IEC 19794.

C.1.2.5 Length of the Head relative to the Image Height

In order to assure that the entire face is visible in the image this part of ISO/IEC 19794 specifies that the crown to chin portion (DD) of the Full Frontal image shall be between 60% and 90% of the vertical length of the image (B) (see clause 8.3.5). Figure 46 shows the distribution of the head height to image height ratio for the four sample data sets.

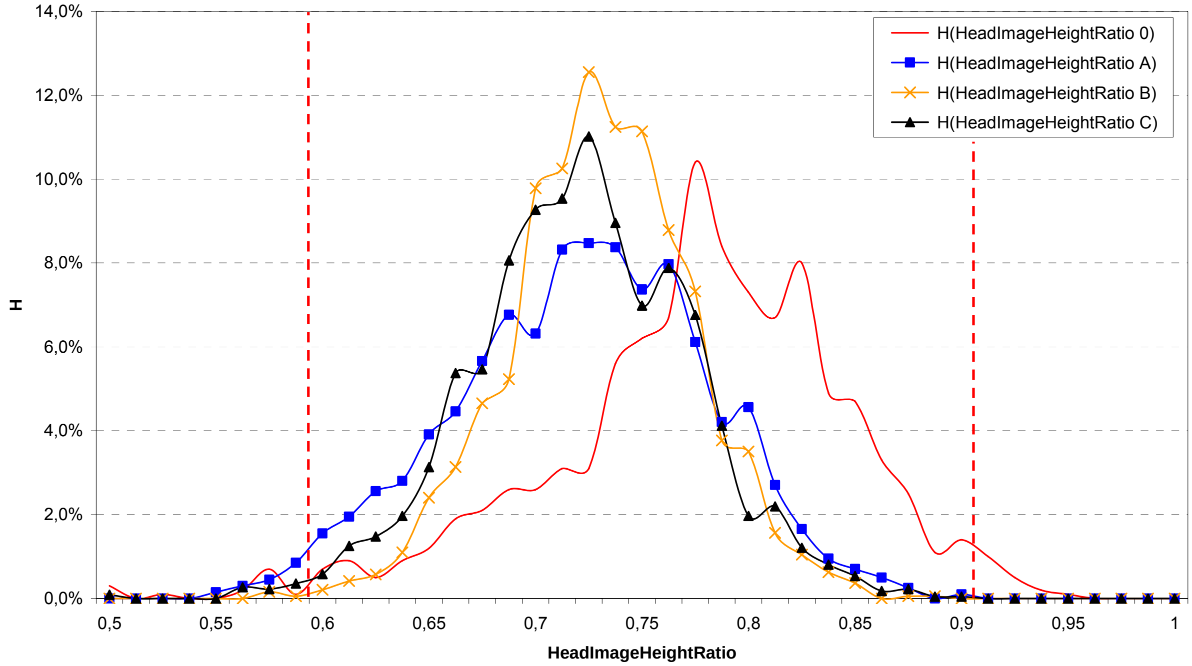


Figure 46 — Distribution of the head to image height ratio in the e-passport images of four Issuing States

The average head to image width ratio of the 7 128 images is found at 0,73. A total of 98.2% of the data fits within the tolerances specified by this part of ISO/IEC 19794.

C.1.3 Error discussion

The analysis of this study is solely based on the measurements done by automated image quality assurance software (QA-SW). No comparison with the so-called ground truth, i.e. the true (manually measured) values for the parameters under consideration, was performed for all the images. However, this was done at an earlier stage for one country's passport images, where the quality assessment software being used was found to be reasonably accurate, and able to produce reliable statistics for larger datasets such as those reported above.

Additional studies comparing the quality assessment software used for this study with other quality assessment software packages on a large number of passport images showed approximately the following deviations:

- QA-SW eye distance: +5%
- QA-SW relative horizontal position: ±1%
- QA-SW relative vertical position: ±1%
- QA-SW head image width ratio: ±1%
- QA-SW head image height ratio: ±1%.

I.e., except for the eye distance, which is reported slightly larger than might be expected to be true, the other parameters may be expected to be correct within an error margin of 1%.

C.1.4 Summary

The study presented in this annex concentrates on geometrical parameters of the face in the photograph which are important for biometric comparison applications. The findings of the study are based on the statistical evaluation of automated image analysis by quality assessment software. Even if individual decisions of the software may have been incorrect, the conclusions based on the aggregation of approximately 7.200 images are certainly valid.

Table 50 — Summary of the compliance of the sample to the requirements specified in this part of ISO/IEC 19794

Criteria	Min	Max	Images compliant
Eye Distance [pixel]	90		99,9%
Relative Horizontal Position	0,45	0,55	95,4%
Relative Vertical Position	0,50	0,70	94,0%
Head to Image Width Ratio	0,50	0,75	98,4%
Head to Image Height Ratio	0,60	0,90	98,2%

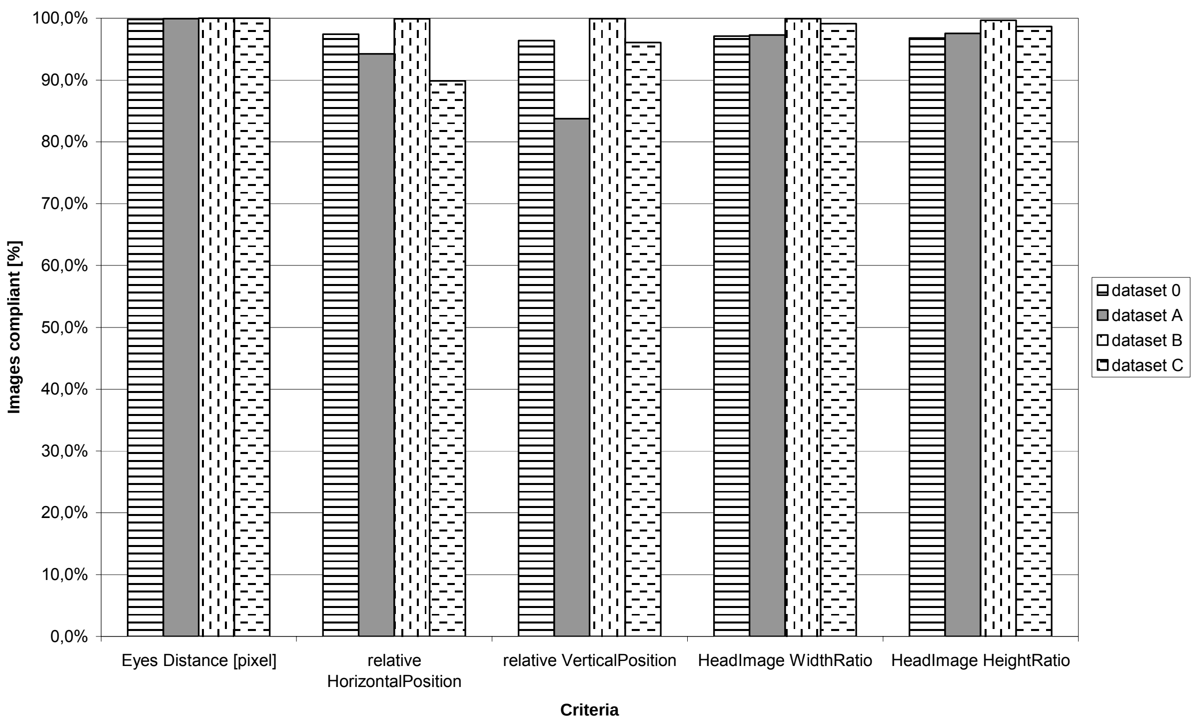


Figure 47 — Percentage of compliant passport photos in the test set of 7 128 images of four Issuing States using the tolerances as put forward in clause 8.3 of this part of ISO/IEC 19794

In summary, the analysis given above shows that with current technology the thresholds set forth in this part of ISO/IEC 19794 can be achieved in real world applications. Table 50 and Figure 47 illustrate a summary of these findings. Since the data used for this study were collected from four major e-passport issuing countries across the world, the results can be regarded as representative within the scope of this analysis.

C.2 Experimental study on the effects of inter-eye distance and the effect of head pose (roll) on biometric matching performance

C.2.1 Inter-eye distance

The distance between the eyes (i.e., the spatial sampling of the photo) is considered to be one of the most important criteria for successfully applying facial recognition. In order to quantify the effect of “pixels between the eyes” on facial recognition performance, images of varying resolution were investigated using a state-of-the-art facial comparison algorithm. Figure 48 shows an increasing verification rate with increasing eye distance. These results also hold for investigations of the one-to-many search performance (rank statistics).

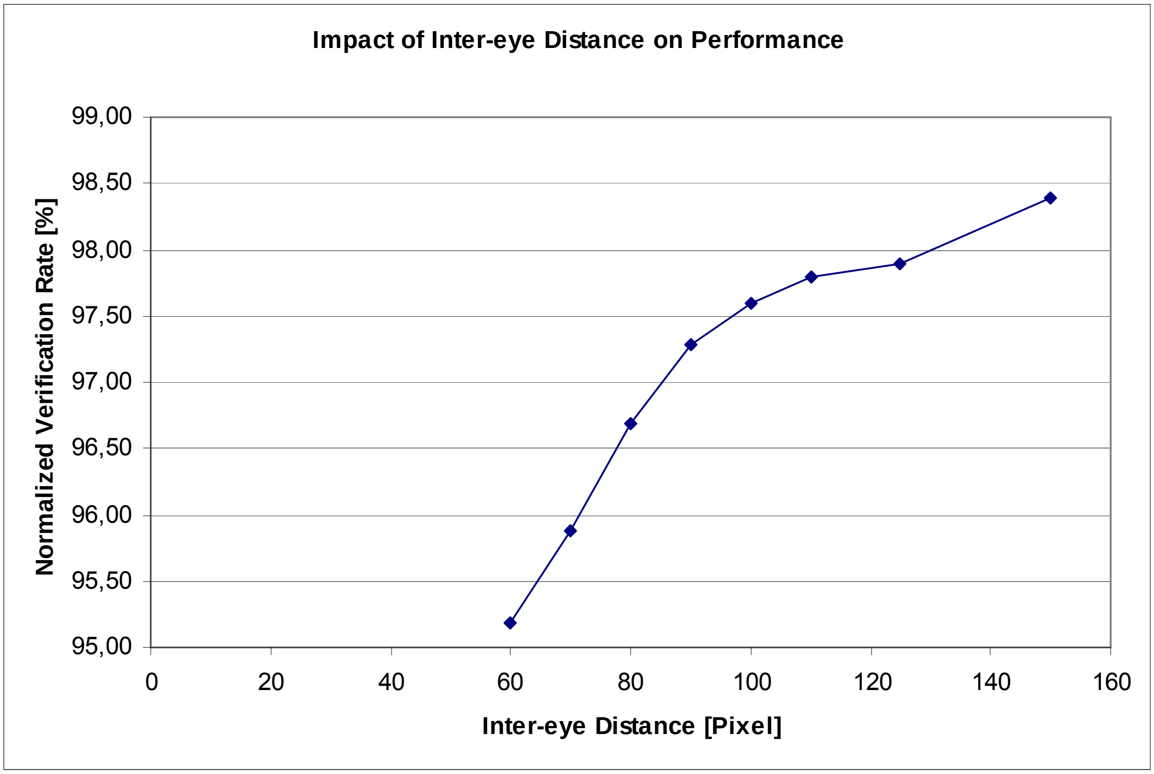


Figure 48 — Normalized verification rate vs. Inter-eye distance at FAR=0,1%

C.2.2 Pose (roll)

Automated face recognition systems allow for effective compensation of the roll angle of a facial image. Figure 49 shows the facial recognition performance impact due to in-plane (roll axis) rotation of the images.

For the test 994 gallery images and 736 probe images from the "Feret Color" dataset have been used. In a first step, all images (probe and gallery) have been rotated, so that they show zero roll angle. Then tests have been carried out, where all probes have been rotated by +5 degrees and all gallery images by -5 degrees, then vice versa. This has been repeated for 10, -10, 15 and -15 degrees, respectively. So, the delta in roll angle of probe and gallery is 0, 10, 20 and 30 degrees. Figure 50 shows some sample data with demonstrating the effect of the rotation.

The verification rate at 0,1% FAR and zero degree has been used to determine the loss due to in-plane rotation, i.e. all other verification rates have been normalized to the verification rate at 0,1% FAR of the zero degree evaluation.

Note, that this setup is the worst case scenario as all images in the ±10 degrees tests show the maximum tolerance defined in this part of ISO/IEC 19794. So, the impact on the performance by a distribution of rotation angles within the maximum tolerances will be significantly less.

Obviously, almost no measurable performance loss can be found up to ± 8 degrees. These results also hold for investigations of the one-to-many search performance (measured by rank1 statistics). Typically, vendors of automated facial recognition software are able to configure the allowed tolerances for in-plane rotation. The software used for these tests has been configured to allow for 10 degrees roll for a single image, i.e. a maximum of 20 degrees difference in the roll angles. This is the reason for the decrease for the ± 15 degrees tests.

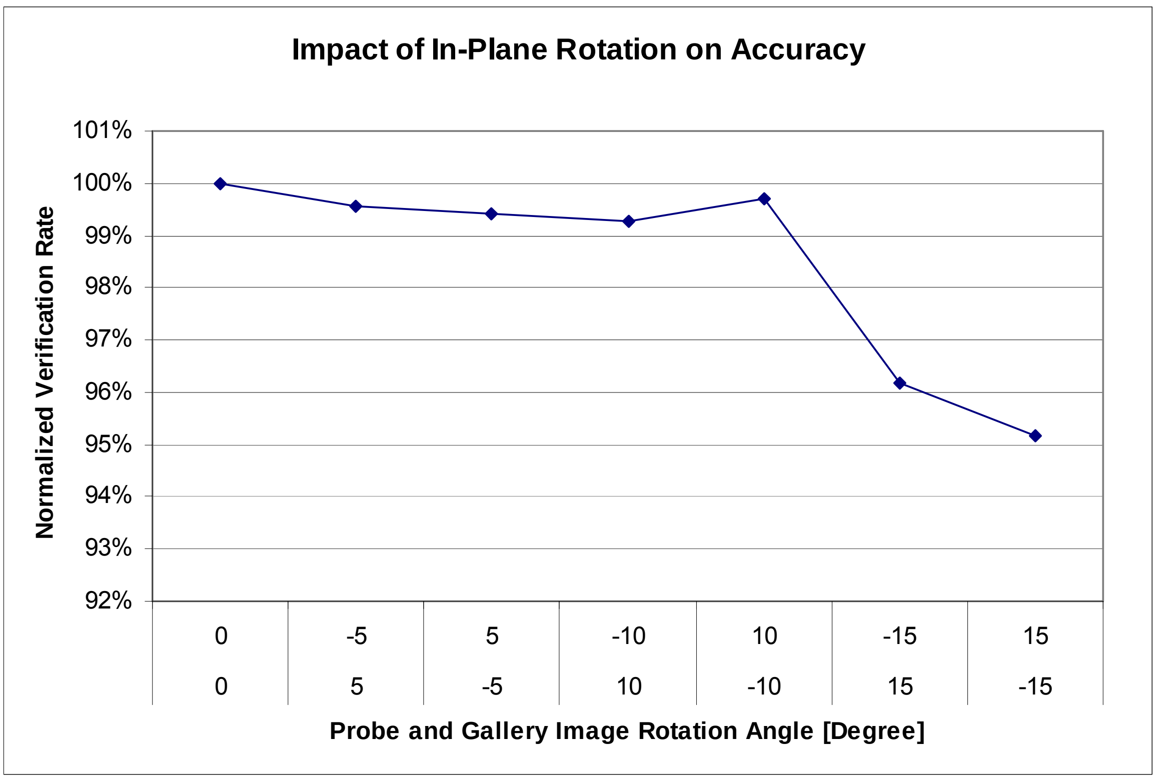


Figure 49 — Impact of in-plane (roll) rotation on automated facial recognition performance: relative verification rate at FAR = 0,1% vs. roll angle

Figure 50 — Suggested Sample images to visualize the in-plane rotation as used in the analysis. The face has a roll angle of 0 degrees, +5 degrees, -5 degrees, +10 degrees, -10 degrees, +15 degrees, -15 degrees (from left to right)

C.2.3 Pose (Pitch and Yaw)

Automated face recognition systems have more difficulty compensating for face pose changes in Pitch and Yaw than changes in Roll. Based on a study performed with 22 subjects and 35 poses per subject differences in Pitch and Yaw greater than the limits set forward in this standard resulted in statistically significant decreases of face recognition performance.

Annex D
(informative)

The Frankfurt Horizon

The Frankfurt horizon is established as a standard plane for orientation of the head. It is defined by a line passing through the right tragion (the front of the ear) and the lowest point of the right eye socket [12], [13].

Note, that the Frankfurt Horizon may be hard to define, as it is related to the ear position that may be covered by hair.

Bibliography

[1] AAMVA DL/ID-2000, American Association of Motor Vehicle Administrators National Standard for the Driver License/Identification Card

[2] ISO/IEC 19784 – Information Technology – The BioAPI Specification

[3] ANSI/NIST-ITL 1-2000, Standard Data Format for the Interchange of Fingerprint, Facial, & Scar Mark & Tattoo (SMT Information)

[4] NISTIR 6322 – Gray Calibration of Digital Cameras To Meet NIST Mugshot Best Practice

[5] NIST Best Practice Recommendation For The Capture Of Mugshots, Version 2.0, 1997

[6] ICC.1:2001-12, File Format for Color Profiles

[7] ISO 10526:1999/CIE S005/E-1998, CIE standard illuminants for colorimetry

[8] ICAO Doc 9303, Part 1 "Machine Readable Passports", 6th edition, 2006

[9] Supplement to ICAO Doc 9303 – Release 6

[10] Anthropometry of the Head and Face, second edition, Leslie G. Farkas, Raven Press, New York, 1994.

[11] The Methods of Plane Projective Geometry Based on the Use of General Homogenous Coordinates, E.A. Maxwell, Cambridge University Press, 1960.

[12] Air Standardization Coordinating Committee (ASCC). (1991). A basis for common practices in the conduct of anthropometric surveys (Air Standard 61/83). Washington, DC

[13] Ranke, J. (ed.). (1884). Verständigung über ein gemeinsames cranio-metrisches Verfahren (Frankfurter Verständigung). Archive Anthropologie, 15, 1-8.

[14] Guides to Quality in Visual Resource Imaging, 3. Imaging Systems: The Range of Factors Affecting Image Quality, Donald P. D'Amato, URL: http://www.rlg.org/visguides/visguide3.html

[15] Commissioned Report: Studies on Biometrics Authentication Using IC Passports, http://www.nmda.or.jp/nmda/bio/ic-passports/ic-passports.pdf

[16] [Special Talk] IC Passports and Systematic Evaluation toward Employing Face Authentication Technologies, Junichi Sakaki, Shizuo Sakamoto, et al., IEICE(The Institute of Electronics, Information and Communication Engineers) technical report PRMU2005-92 (2005-10), pp. 51-56, 2005 (in Japanese).

[17] FRVT 2006 and ICE 2006 Large-Scale Results, P. Jonathon Phillips, et al., NISTIR 7408, National Institute of Standards and Technology, Gaithersburg, MD 20899, USA, URL: http://face.nist.gov/frvt/frvt2006/frvt2006.htm

[18] C-Cube Microsystems, JPEG File Interchange Format (JFIF), Version 1.02

[19] PIMA 7667:2001, Photography — Electronic Still Picture Imaging — Extended sRGB Color Encoding — e-sRGB

