INTERNATIONAL STANDARD

ISO/IEC 19794-5

First edition
2005-06-15

Corrected version
2016-09-01

Information technology — Biometric data interchange formats —

Part 5:
Face image data

Technologies de l'information — Formats d'échange de données biométriques —

Partie 5: Données d'image de la face

Reference number
ISO/IEC 19794-5:2005(E)

Contents

Page

Foreword viii
Introduction ix
1 Scope 1
2 Compliance 2
3 Normative references 2
4 Terms and definitions 3
5 The Face Image Record Format 6
5.1 Overview 6
5.2 Data Conventions 9
5.2.1 Byte ordering 9
5.2.2 Numeric values 9
5.2.3 Conversion to integer 9
5.2.4 Unspecified field value 9
5.2.5 Unknown field value 9
5.3 The CBEFF Header 9
5.4 The Facial Record Header 10
5.4.1 Format Identifier 10
5.4.2 Version Number 10
5.4.3 Length of Record 10
5.4.4 Number of Facial Images 10
5.5 The Facial Information Block 10
5.5.1 Facial Record Data Length 11
5.5.2 Number of Feature Points 11
5.5.3 Gender 11
5.5.4 Eye Colour 11
5.5.5 Hair Colour 12
5.5.6 Property Mask 12
5.5.7 Expression 13
5.5.8 Pose Angle 13
5.5.9 Pose Angle Uncertainty 15
5.6 The Landmark Point Block 15
5.6.1 Landmark Point Type 16
5.6.2 Landmark Point Code 16
5.6.3 MPEG4 Feature Points 16
5.6.4 Eye and nostril Landmark Points 17
5.6.5 Anthropometric Landmarks 18
5.6.6 Anthropometric 3D landmark 21
5.6.7 Z Coordinate 21
5.7 The Image Information Block 22
5.7.1 Face Image Type 22
5.7.2 Image Data Type 23
5.7.3 Width 23
5.7.4 Height 23
5.7.5 Image Colour Space 23
5.7.6 Source Type 23
5.7.7 Device Type 24
5.7.8 Quality 24
5.8 The Image Data Block 24
5.8.1 Data structure 24
5.9 The 3D Information Block 24

5.9.1 Length of 3D Data Representation 25
5.9.2 Coordinate System Type 25
5.9.3 Texture Projection Matrix 27
5.9.4 ScaleX, ScaleY, ScaleZ, OffsetX, OffsetY, OffsetZ 27
5.9.5 3D Representation Type 28
5.9.6 3D Supplemental Data 28
5.9.7 3D Source Type 28
5.9.8 3D Device Type 29
5.9.9 3D to 2D Image Temporal Synchronicity 29
5.9.10 3D to 2D Texture Temporal Synchronicity 29
5.9.11 3D Acquisition Time 30
5.9.12 2D Texture Acquisition Time 30
5.9.13 Texture Map Type 30
5.9.14 Texture Map Spectrum 31
5.10 The 3D Data Block 31
5.10.1 Range Image Bit Depth 31
5.10.2 Range Image 32
5.10.3 3D Point Map Width and Height 32
5.10.4 3D Point Map 32
5.10.5 Vertex Data 32
5.10.6 Triangle Data 33
5.10.7 Error Map 33
5.10.8 Texture Map 33

6 The Basic Face Image Type 34
6.1 Inheritance requirements for the Basic Face Image Type 34
6.2 Image data encoding requirements for the Basic Face Image Type 34
6.3 Image data compression requirements for the Basic Face Image Type 34
6.4 Format requirements for the Basic Face Image Type 34
6.4.1 Facial Header 34
6.4.2 Facial Information 34
6.4.3 Image Information 34

7 The Frontal Face Image Type 34
7.1 Inheritance requirements for the Frontal Face Image Type 34
7.2 Scene requirements for the Frontal Image Type 35
7.2.1 Purpose 35
7.2.2 Pose 35
7.2.3 Expression 35
7.2.4 Assistance in positioning the face 36
7.2.5 Shoulders 36
7.2.6 Backgrounds 36
7.2.7 Subject and scene lighting 36
7.2.8 Shadows over the face 36
7.2.9 Shadows in eye sockets 36
7.2.10 Hot spots 36
7.2.11 Eye glasses 36
7.2.12 Eye patches 36
7.3 Photographic Requirements for the Frontal Image Type 37
7.3.1 Purpose 37
7.3.2 No over or under exposure 37
7.3.3 Focus and depth of field 37
7.3.4 Unnatural colour 37
7.3.5 Colour or greyscale enhancement 37
7.3.6 Radial distortion of the camera lens 37
7.4 Digital requirements for the Frontal Image Type 37
7.4.1 Geometry 37
7.4.2 Colour profile 38
7.4.3 Video interlacing 38
7.5 Format requirements for the Frontal Image Type 38
7.5.1 Inheritance requirements 38

7.5.2 Image Information 38

8 The Full Frontal Image Type 39
8.1 Inheritance requirements for the Full Frontal Face Image Type 39
8.2 Scene requirements for the Full Frontal Face Image Type 39
8.3 Photographic requirements for the Full Frontal Face Image Type 39
8.3.1 Introduction 39
8.3.2 Horizontally centred face 40
8.3.3 Vertical position of the face 40
8.3.4 Width of head 40
8.3.5 Length of head 40
8.3.6 Summary of photographic requirements 40
8.4 Digital requirements for the Full Frontal Face Image Type 41
8.4.1 Resolution 41
8.5 Format requirements for the Full Frontal Image Type 41
8.5.1 Inheritance requirements 41
8.5.2 Image Information 41

9 The Token Face Image Type 41
9.1 Inheritance requirements for Token Face Image Type 41
9.2 Digital requirements for the Token Face Image Type 42
9.2.1 Introduction 42
9.2.2 Eye positions 42
9.2.3 Token image geometric format 42
9.2.4 Minimum width Token image 43
9.2.5 Padding 43
9.3 Format requirements for the Token Face Image Type 43
9.3.1 Inheritance requirements 43
9.3.2 Image Information 43

10. The Basic 3D Image Type 43
10.1 Inheritance Requirements for the Basic 3D Image Type 43
10.2 The Basic 3D Image Type using the 3D Point Map representation 44
10.2.1 Coordinate System Type 44
10.2.2 ScaleX, ScaleY and ScaleZ 44
10.3 The Basic 3D Image Type using the 3D Vertex representation 44
10.3.1 Coordinate System Type 44
10.3.2 ScaleX, ScaleY and ScaleZ 44

11 The Full Frontal 3D Image Type 44
11.1 Inheritance requirements 44
11.2 Coordinate System Type 44
11.3 Pose of the 3D representation 44
11.4 Calibration Texture Projection Accuracy 45
11.5 Requirements on Full Frontal 3D Image Types using the Range Image Representation 45
11.5.1 ScaleX, ScaleY and ScaleZ 45
11.5.2 Face Coverage 45
11.5.3 Non-valid points in 3D data Image 45
11.6 Requirements on Full Frontal 3D Image Types using the 3D Point Map Representation 46
11.7 Requirements on Full Frontal 3D Image Types using the 3D Vertex Representation 46

12 The Token Frontal 3D Image Type 46
12.1 General 46
12.2 Inheritance requirements 46
12.3 Requirements on Token Frontal 3D Image Types using the Range Image Representation 47
12.4 Requirements on Token Frontal 3D Image Types using the 3D Point Map Representation 47
12.5 Requirements on Token Frontal 3D Image Types using the Vertex Representation 47

Bibliography 48

Annex A 49
A.1 Best practices for Basic Face Images 49
A.1.1 Purpose 49
A.1.2 Feature Point determination 49

A.2 Best practices for Frontal Images 49
A.2.1 Purpose 49
A.2.2 Pose 49
A.2.3 Expression 49
A.2.4 Assistance in positioning the face 49
A.2.5 Background 50
A.2.5.1 Background segmentation 50
A.2.5.2 Background shadows 50
A.2.5.3 Background uniformity 50
A.2.5.4 Background examples 50
A.2.6 Focus and depth of field 50
A.2.7 No unnatural colour 50
A.2.8 Colour calibration 50
A.2.9 Radial distortion of the camera lens 50
A.3 Best practices for Full Frontal Images 51
A.3.1 Digital attributes of Full Frontal Images 51
A.3.1.1 Photo resolution 51
A.3.2 Best practices for use of Full Frontal Images on Travel Documents 51
A.3.2.1 Width to height ratio of the image 51
A.3.2.2 Head size relative to the image size 51
A.3.2.3 Summary of best practice photographic recommendations 51
A.3.2.4 Sample images and sample photograph taking guidelines for travel documents 53
A.3.3 Full Frontal Image compression 56
A.3.3.1 Compression – no region of interest 56
A.3.3.2 Recommendations for maximum compression and file sizes for JPEG and JPEG2000 57
A.3.4 Full Frontal Image compression using region of interest 57
A.3.4.1 Discussion 57
A.3.4.2 Inner and outer regions, Full Image 58
A.4 Best practices for Token Images 58
A.4.1 Token image sizes 58
A.4.2 Creation of a Token Image 59
A.4.3 Best practices for digital attributes of Token Images 59
A.4.4 Token Image compression 60
A.4.4.1 Compression – no region of interest 60
A.4.4.2 Recommendations for maximum compression and file sizes for JPEG and JPEG2000 Token Images 61
A.4.5 Token Image compression using region of interest 61
A.4.5.1 Discussion 61
A.4.6 Inner and outer regions for the Token Image for the purpose of compression 62
A.5 Experimental study on the enrolment of full frontal images for travel documents 62
A.5.1 Software and data used for the analysis 62
A.5.2 Experimental results 63
A.5.2.1 Inter-eye distance 63
A.5.2.2 Relative horizontal position of the face 64
A.5.2.3 Relative vertical position of the face 64
A.5.2.4 Head Image Width Ratio 65
A.5.2.5 Head Image Height Ratio 66
A.5.3 Error Discussion 67
A.5.4 Summary 67
A.6 Study on the effects of inter-eye distance and roll on biometric comparison performance 68
A.6.1 Inter-eye distance 68
A.6.2 Pose 69
A.7 Best Practices for the Full Frontal 3D Image Type 70
A.7.1 Best Practices for the 2D part of the Full Frontal 3D Image Type 70
A.7.2 Compatibility considerations 70
A.7.3 Pose of the 3D representation 70
A.7.4 3D to 2D Image Temporal Synchronicity 71
A.7.5 3D Acquisition Time 71
A.7.6 Best Practices for Full Frontal 3D Image Types using the Range Image Representation 71
A.7.6.1 ScaleX, ScaleY and ScaleZ 71

A.7.6.2 Non-valid points in Range Image 71
A.7.7 Best Practices for the Full Frontal 3D Image Types using the 3D Point Map Representation 71
A.7.7.1 3D Point Map Width and Height 71
A.7.7.2 Face coverage 71
A.7.8 Best Practices for Full Frontal 3D Image Types using the 3D Vertex Representation 71
A.7.8.1 Face coverage 71
A.8 Best Practices for Token Frontal 3D Images 72
A.8.1 Best Practices for the 2D part of the Token Frontal 3D Image 72
A.8.2 Compatibility considerations 72
A.8.3 Pose of the 3D representation 72
A.8.4 3D to 2D Image Temporal Synchronicity 72
A.8.5 3D Acquisition Time 72
A.8.6 Best Practices for Token Frontal 3D Image Types using the Range Image Representation 72
A.8.7 Best Practices for Token Frontal 3D Image Types using the 3D Point Map Image Representation 72
A.8.8 Best Practices for Token Frontal 3D Image Types using the Vertex Representation 72
A.9 Summary of mandatory and best practices for the 3D Image Types 72

Annex B 75
B.1 Scope 75
B.2 Photography recommendations 75
B.2.1 General 75
B.2.2 Recommendations for a photo studio or store 75
B.2.3 Recommendations for photo booths 79
B.2.4 Recommendations for a registration office environment 83
B.3 Guidelines for printing 84
B.3.1 General 84
B.3.2 Spatial and tonal resolution trade-offs 85
B.3.3 Recommended printing quality 85
B.3.4 Use of a photo template 86
B.4 Guidelines for scanning 86
B.4.1 General 86
B.4.2 Sampling frequency and quantization levels 87
B.4.3 Spatial resolution 87
B.4.4 Output colour space 87
B.4.5 Saturation 87
B.4.6 Image compression 87
B.5 Face image quality assessment software 87
B.6 Tables of the recommendations 89
B.6.1 General 89
B.6.2 Scene setting 89
B.6.3 Photographing 91
B.6.4 After photographing 91
B.6.5 Photographic quality 92
B.7 Experimental data 93
B.7.1 Experimental results of face recognition in a photo studio and photo booth 93
B.8 Photographic examples 94
B.8.1 General 94
B.8.2 Photographic examples at a photo studio 94
B.8.3 Photographic examples at a photo booth 99

Annex C 104

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National Bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

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

The main task of the joint technical committee is to prepare International Standards. Draft International Standards adopted by the joint technical committee are circulated to 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 and IEC shall not be held responsible for identifying any or all such patent rights.

This consolidated reprint of ISO/IEC 19794-5:2005 was prepared by Joint Technical Committee ISO/IEC JTC1, *Information Technology*, Subcommittee SC 37, *Biometrics*.

This consolidated reprint of ISO/IEC 19794-5:2005 contains the original content of ISO/IEC 19794-5:2005 and incorporates the Amendments ISO/IEC 19794-5:2005/Amd 1:2007 and ISO/IEC 19794-5:2005/Amd 2:2009 and the Technical Corrigenda ISO/IEC 19794-5:2005/Cor 1:2008, ISO/IEC 19794-5:2005/Cor 2:2008, ISO/IEC 19794-5:2005/Cor 3:2013, ISO/IEC 19794-5:2005/Cor 4:2015, and the unpublished draft of ISO/IEC 19794-5:2005/Cor 5.

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 document is intended to provide a face image format for face recognition applications requiring exchange of face image data. The typical applications are:

- 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;
- Human verification of identify by comparison of persons against facial images;
- Computer automated face identification (one-to-many searching);
- Computer automated face verification (one-to-one matching).

To enable many applications on variety of devices, including devices that have the limited resources required for data storage, and to improve face recognition accuracy, this document 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.

The record format specified in this document is designed to be embedded in a CBEFF-compliant structure specified in the multi-part Standard ISO/IEC 19785. The embedment in the CBEFF structure is described in ISO/IEC 19794-1:2006.

Information technology – Biometric data interchange formats – Face image data

1 Scope

This document specifies

- A record format for storing, recording, and transmitting the information from one or more facial images within a CBEFF data structure,
- Scene constraints of the facial images,
- Photographic properties of the facial images,
- Digital image attributes of the facial images.

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 image 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 "mug shot" 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.

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	9	A.4

Figure 1 gives a general overview of the scene, photographic, digitization, and format requirements for the face image types specified in this document. The *Basic Face Image Type* has no scene, photographic, or digital requirements.

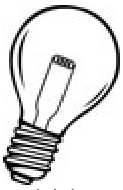
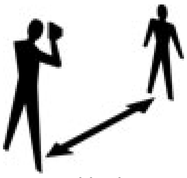
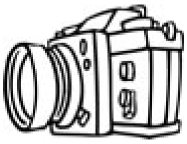
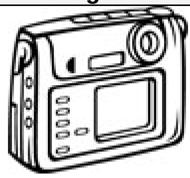
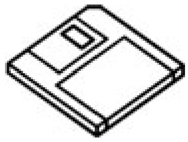
Requirements			
Scene	**Photographic**	**Digital**	**Format**
 Lighting  Image and Subject	 Positioning  Camera Attributes	 Digital Camera  Analogue to Digital  Image Scanning	 Digital Specifications  Record Format and Organization
Clauses: Basic Face None Frontal Face 7.2 Full Frontal Face 8.2	*Clauses:* 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

Figure 1 – Types of imaging requirements specified in this document.

2 Compliance

Conformity with this document requires compliance with the record format specification defined in Clauses 5 and the *Basic Face Image Type* defined in Clause 6. In addition, this document defines additional *Face Image Types*. Compliance with the *Full Frontal Face Image Type* requires compliance with Clauses 5, 6, 7, and 8. Conformity with the *Token Frontal Image Type* requires additional compliance with Clause 5, 6, 7, and 9.

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 19785 Information technology -- Common Biometric Exchange Formats Framework – Multi-part Standard
- ISO/IEC 19794-1:2006 Information technology — Biometric data interchange formats — Part 1: Framework

- ISO/IEC 10918-1, Information technology — Digital compression and coding of continuous-tone still images: Requirements and guidelines
- ISO/IEC 15444-1, Information technology — JPEG 2000 image coding system: Core coding system
- C-Cube Microsystems, JPEG File Interchange Format (JFIF), Version 1.02
- PIMA 7667:2001, Photography – Electronic Still Picture Imaging – Extended sRGB Color Encoding – e-sRGB
- ICC.1:2001-12, File Format for Color Profiles
- ISO/IEC 14496-2:2004, Information technology - Coding of audio-visual objects - Part 2: Visual (MPEG4)
- ISO/IEC 15948:2004, Information technology — Computer graphics and image processing — Portable Network Graphics (PNG): Functional specification

4 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 19794-1:2006 and the following apply.

4.1
chin
The central forward portion of the lower jaw.

4.2
colour image
Continuous tone image that has more than one channel, each of which is coded with one or multiple bits.

4.3
colour space
A way of representing colours of pixels in an image. For instance, RGB, YUV and greyscale are typically used in this document.

4.4
common biometric exchange formats framework, CBEFF
Data format specifically for exchanging biometric data that provides for the encompassing of any biometric type into a standard format.

4.5
continuous tone image
Image whose channels have more than one bit per pixel.

4.6
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.

4.7
dots per inch, DPI, dpi
Measurement of scanner and printer resolution.

4.8
facial image
Electronic image-based representation of the portrait of a person.

4.9
Face Image Type
A category of facial images that satisfy specific requirements.

4.10
FAP
Facial Animation Parameter.

4.11
fish eye
A type of distortion where central objects of the image erroneously appear closer than those at the edge.

4.12
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.

4.13
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.

4.14
human verification
Process of human comparison of a face image with a person or another face image to ascertain the identity of the respective person in a short time period; one-to-one (1:1) matching.

4.15
identification
Process of searching through a list of face images to match against an input image(s); one-to-many (1:N) searching.

4.16
2D image
Two-dimensional representation that encodes the luminance and/or colour texture of a capture subject in a given lighting environment.

4.17
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.

4.18
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.

4.19
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.

4.20
pixel
Picture element; element on a two-dimensional array that comprises an image.

4.21
portrait
Photograph of a person which includes the full head, with all hair in most cases, as well as neck and top of shoulders.

4.22
red-eye
The red glow from subject's eye caused by light from flash reflecting from blood vessels behind the retina.

4.23
verification
Process of ascertaining that two images or image inputs represent the same person; one-to-one (1:1) comparison.

4.24
3D image
Representation that encodes a surface in a 3D space.

4.25
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.

4.26
3D vertex representation
Representation using 3D vertices and triangles between these points for coding of a 3D surface.

4.27
anthropometric landmark
Landmark point on the face used for identification and classification of humans.

4.28
anthropometric landmark code
Two-part code that defines an anthropometric landmark uniquely.

4.29
cartesian coordinate system
3D orthogonal coordinate system.

4.30
cylindrical coordinate system
Three-dimensional polar coordinate system describing a point by the three components radius, azimuth and height.

4.31
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.

4.32
PNG format
Lossless image compression standard specified in ISO/IEC 15948.

4.33
texture
Two-dimensional representation of the luminance and/or colour of a capture subject in a given lighting environment

4.34
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 reference [11] for details.

5 The Face Image Record Format

5.1 Overview

The *Face Image Record Format* specified in this document is a format to store face image data within a biometric data record. Each record 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 image. This record is embedded in the biometric data block in a CBEFF compliant structure. 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, the header and the entire data structure will be CBEFF compatible, 2D image data will be encoded using either JPEG or JPEG2000. The 3D Information block and 3D Data block are mandatory for the 3D Types.

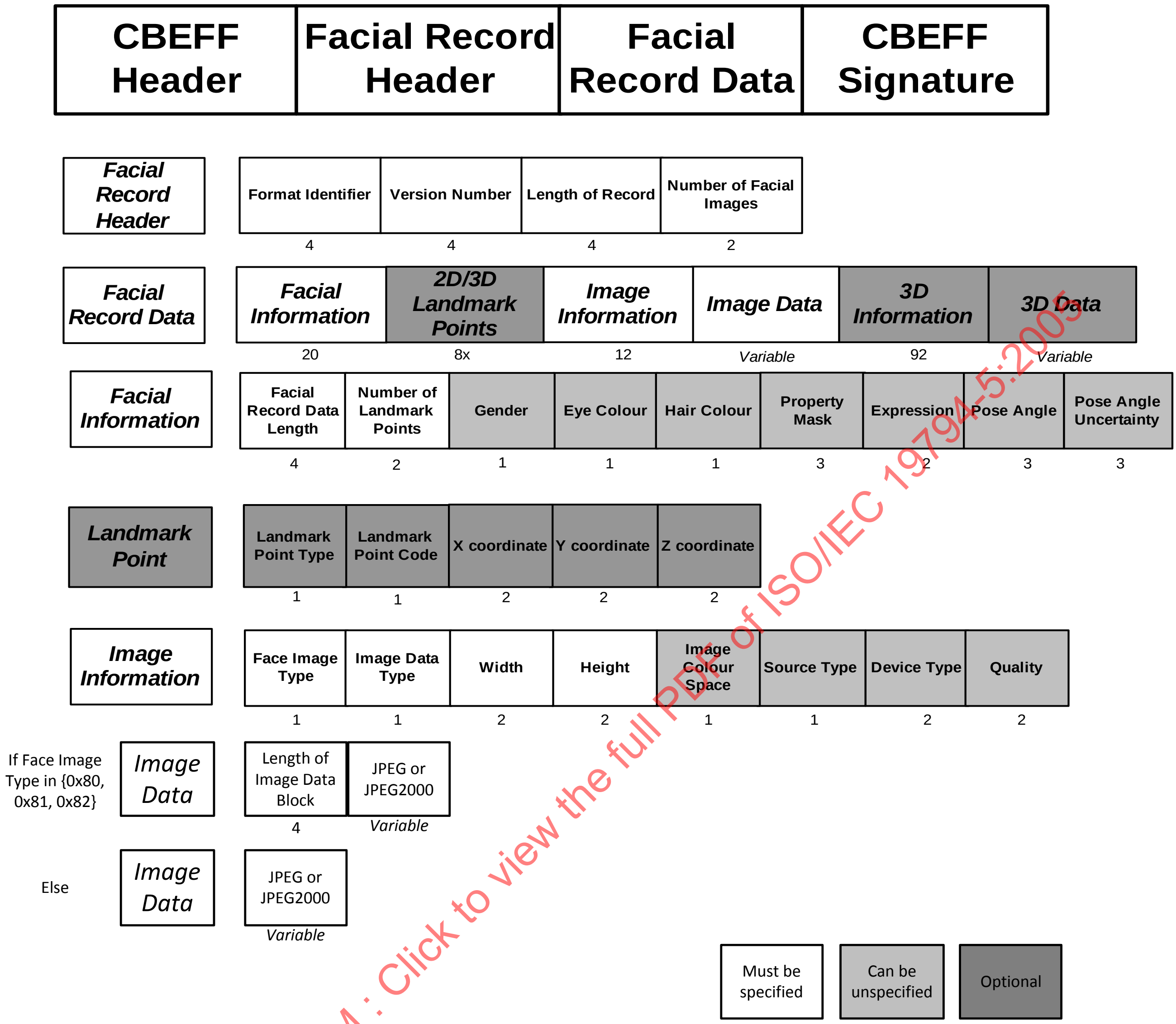


Figure 2 — *Face Image Record Format*. The length value of each field in bytes is shown below the field. The white boxes indicate fields or blocks that shall be specified, light grey boxes that the fields are mandatory, but an unspecified value is acceptable, and dark grey boxes indicate optional fields. The *Length of Image Data Block* element is mandatory if the *Face Image Type* as described in Clause 5.7.1 is a 3D type (0x80, 0x81, 0x82). For all other *Face Image Types* there must not be a *Length of Image Data Block* element.

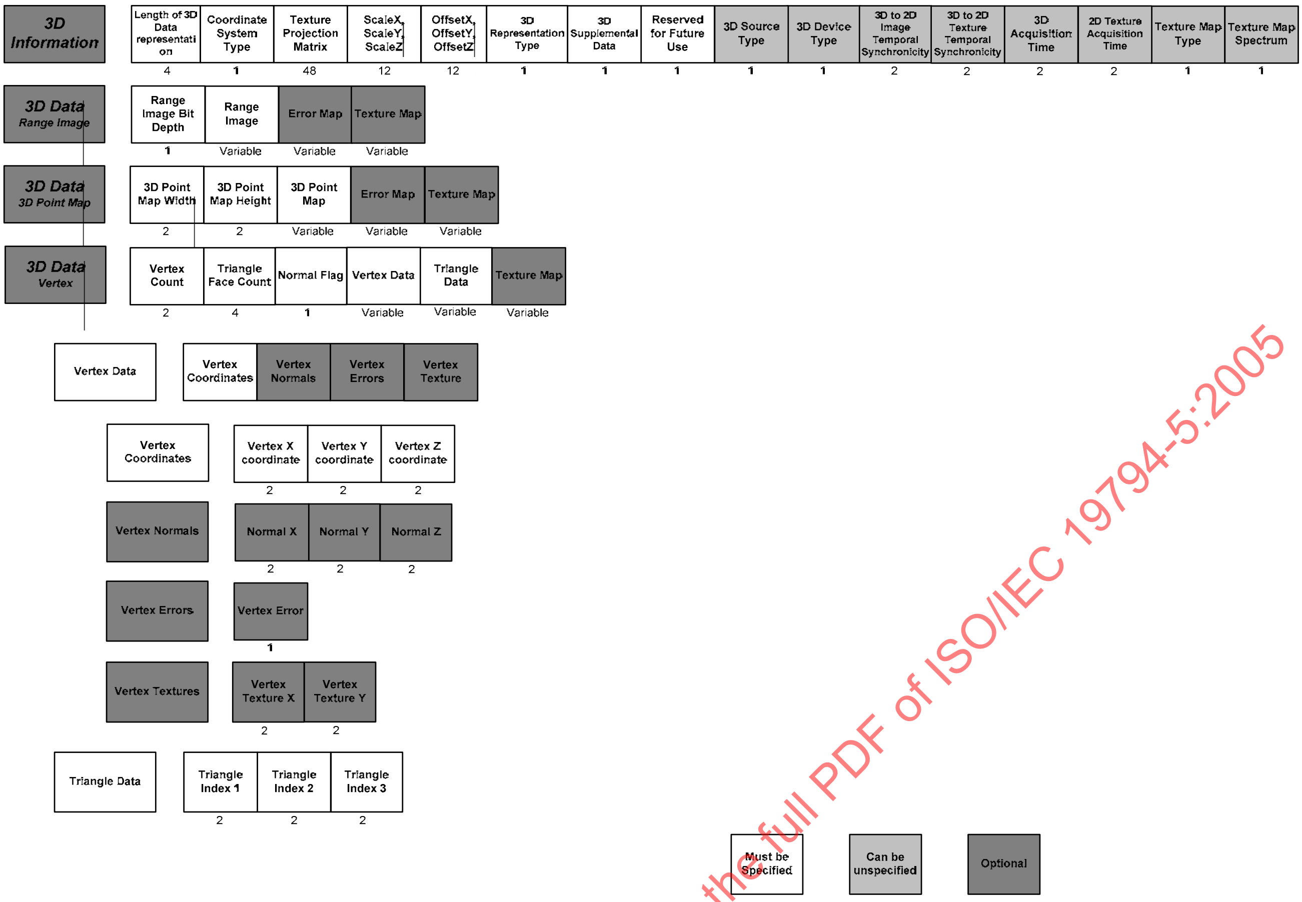


Figure 3 — *3D Information Block* and the three possible *3D Data Blocks*. The length value of each field in bytes is shown below the field. The white boxes indicate fields or blocks that shall be specified, light grey boxes that the fields are mandatory, but an unspecified value is acceptable, and dark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

When referring to elements of the record format, *field* denotes the elementary unit of information 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 biometric reference which consists of the *Facial Record Header* and one or more *Facial Record Data*. With the exception of the *Format Identifier* and the *Version Number*, 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 (14 byte) *Facial Record Header* containing information about the overall record, including the number of facial images represented and the overall record length in bytes;
- A *Facial Record Data Block* for each facial image. This data consists of
  - A fixed length (20 byte) *Facial Information Block* describing discernible characteristics of the subject such as gender.
  - Multiple (including none) fixed length (8 byte) *Landmark Point Blocks* describing 2D or 3D *Landmark Points* on a face.
  - A fixed length (12 byte) *Image Information Block* describing digital properties of the image such as *Face Image Type* and dimensions such as width and height.
  - *Image Data* consisting of a JPEG or JPEG2000 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 images/3D-representations of the same biometric data subject can be described in a single CBEFF record. This is accomplished by including multiple *Facial Record Data Blocks* after the *Facial Record Header Block* and before the CBEFF signature block. *Facial Record Data Blocks* containing 2D data can be stored together with *Facial Record Data Blocks* also containing 3D data. The structure of this embedding is illustrated in Figure 4.

CBEFF Header	Facial Record Header	Facial Record Data 1 (Full Frontal)	Facial Record Data 2 (Token Frontal)	Facial Record Data 3 (Full Frontal Range)	CBEFF Signature

Figure 4 – Embedding multiple images/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 1025 (2 to the 10th 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

The field value of zero (0x00) shall be used to denote that the creator of the record did not make the determination of the information encoded in the field. The only exception is the value of zero assigned to JPEG in the *Image Data Type* field in Clause 5.7.2.

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 CBEFF Header

The biometric data record represented using the face record format shall be embedded in the biometric data block (BDB) of the CBEFF patron format compliant with ISO/IEC 19785-1:2004. The CBEFF patron format requests to specify both *CBEFF_BDB_format_owner* and *CBEFF_BDB_format_type* as mandatory elements in the CBEFF Header. The *CBEFF_BDB_format_owner* shall be specified by the CBEFF biometric organization identifier issued by the CBEFF registration authority to ISO/IEC JTC1/SC37. This value is the sixteen bit value 0x0101. The *CBEFF_BDB_format_type* shall be specified by the CBEFF BDB format type identifier assigned by ISO/IEC JTC1/SC37 to this face record format. This value is the sixteen bit value 0x0008. Complete CBEFF header information required for coding is given in ISO/IEC 19794-1:2006.

5.4 The Facial Record Header

The *Facial Record Header* block consists of four fields; *Format Identifier*, *Version Number*, *Length of Record*, *Number of Facial Images* as shown in Table 2.

Table 2 – Facial Record Header

Field	Size	Valid values	Notes
Format Identifier	4 bytes	0x46414300 (‘F’ ‘A’ ‘C’ 0x0)	Indicates face image data
Version Number	4 bytes	0x30313000 (‘0’ ’ 1’ ’0’ 0x0)	“010” in ASCII
Length of Record	4 bytes	$57 < \text{Length of Record} \leq 2^{32} - 1$	Includes Facial Record Header and Facial Record Data. The minimum length of 57 includes the smallest JPEG image.
Number of Facial Images / 3D representations	2 bytes	$1 \leq \text{Number} \leq 65535$	

5.4.1 Format Identifier

The (4 byte) *Format Identifier* shall consist of three ASCII characters “*FAC*” followed by a zero byte as a NULL string terminator to identify the record format as the face record format.

5.4.2 Version Number

The (4 byte) *Version Number* field shall consist of three ASCII numerals followed by a zero byte as a NULL string terminator. The first and second characters represent the major version number and the third character represents the minor revision number.

The *Version Number* shall be 0x30313000; “010” – Version 1 revision 0.

The *Version Number* in some implementations may also be 0x30323000, that is, “020”, however, this setting is deprecated.

5.4.3 Length of Record

The (4 byte) *Length of Record* field shall be the combined length in bytes for the record. This is the entire length of the record including the *Facial Record Header* and *Facial Record Data*.

5.4.4 Number of Facial Images

The (2 byte) *Number of Facial Images* field shall be the number of facial images included in the record.

5.5 The Facial Information Block

The (20 byte) *Facial Information Block* is intended to describe discrete properties of the individual discernible from the image; one is included for each facial image included in the record. The structure of this block is shown in Figure 2. Zero or more *Feature Point Blocks*, one *Image Information Block*, and one *Image Data Block* follow this block.

5.5.1 Facial Record Data Length

The (4 byte) *Facial Record Data Length* field denotes the sum of the lengths of the *Facial Information Block*, the *Feature Point Block(s)*, the *Image Information Block*, and the *Image Data Block*. The minimum value of the *Facial Record Data Length* is 32 bytes plus the size of the *Image Data Block* (in bytes).

5.5.2 Number of Feature Points

The (2 byte) *Number of Feature Points* field shall be the number of *Feature Point Blocks* that follow the *Facial Information Block*. The *Feature Point Block* is defined in Clause 5.6.

5.5.3 Gender

The (1 byte) *Gender* field shall represent the gender of the subject according to Table 3.

Table 3 – Gender codes

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

5.5.4 Eye Colour

The (1 byte) *Eye Colour* field shall represent the colour of irises of the eyes according to Table 4. If the eyes have different colours, then the colour of the right eye has to be encoded.

Table 4 – Eye Colour codes

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

5.5.5 Hair Colour

The (1 byte) *Hair Colour* field shall represent the colour of the hair according to the Table 5.

Table 5 – Hair Colour codes

Description	Value
Unspecified	0x00
Bald	0x01
Black	0x02
Blonde	0x03
Brown	0x04
Gray	0x05
White	0x06
Red	0x07
Reserved	0x08 – 0xFE
Unknown or Other	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 6 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 (and all bits shall be zero); 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. All reserved bits shall be zero.

Table 6 – Property flags

Description	Mask Position	Description	Mask Position
Properties are specified	0	Mouth open	6
Glasses	1	Left eye patch	7
Moustache	2	Right eye patch	8
Beard	3	Dark glasses (medical)	9
Teeth visible	4	Feature distorting medical condition (which could impact feature point detection)	10
Blink (either or both eyes closed)	5	Reserved	11 – 23

A Blink flag set to “1” shall 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 7.

Table 7 – Expression codes

Description	High Byte	Low Byte
Unspecified	0x00	0x00
Neutral (non-smiling) with both eyes open and mouth closed)	0x00	0x01
A smile where the inside of the mouth and/or teeth is not exposed (closed jaw).	0x00	0x02
A smile where the inside of the mouth and/or teeth is exposed.	0x00	0x03
Raised eyebrows	0x00	0x04
Eyes looking away from the camera	0x00	0x05
Squinting	0x00	0x06
Frowning	0x00	0x07
Reserved	0x00 0x01 – 0x7F	0x08 – 0xFF 0x00 – 0xFF
Vendor Specific	0x80 – 0xFF	0x00 – 0xFF

5.5.8 Pose Angle

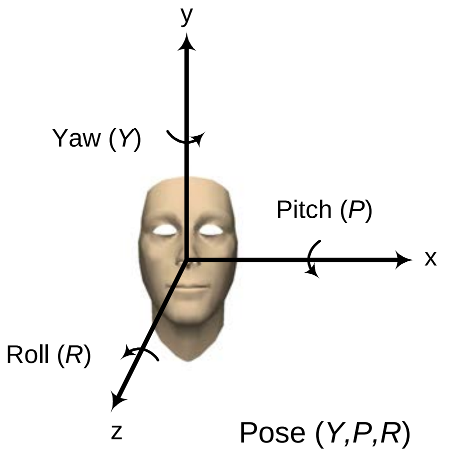


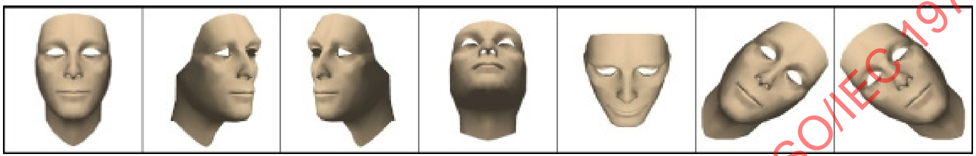
Figure 5 – The definition of pose angles is with respect to the frontal view of the subject.

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 view of the subject, which has angles (0, 0, 0) as shown in Figure 5. 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: 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.

For example, for the facial images in Figure 6 the pose angles (Y, P, R) of sub-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* faces have a yaw angle of 0 degrees. Positive angles represent faces looking to their left (a counter-clockwise rotation around the y-axis). The encoded value, B_Y, shall be stored in 1 byte with values 0 to 180 computed from a real-valued yaw angle estimate, $-180 \le Y < 180$, as follows:

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

The maximum value of B_Y is 180. 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* faces have a pitch angle of 0 degrees. Positive angles represent faces looking down (a counter-clockwise rotation around the x-axis). The encoded value, B_P, shall be stored in 1 byte with values 0 to 180 computed from a real-valued pitch angle estimate, $-180 \le P < 180$, as follows:

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

The maximum value of B_P is 180. If the pitch angle is not specified, the value of B_P shall be 0.

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 faces have a roll angle of 0 degrees. Positive angles represent faces tilted toward their right shoulder (counter-clockwise rotation around the z-axis). The encoded value, B_R, shall be stored in 1 byte with values 0 to 180 computed from a real-valued roll angle estimate, $-180 \le R < 180$, as follows:

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

The maximum value of B_R is 180. 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 angles 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

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 8. *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 and their relation with the set of MPEG4 feature points is discussed in Clause 5.6.5 of this document. The horizontal and vertical positions of *Landmark Points* are either texture image coordinates or in the Cartesian coordinate system (see Clause 5.9.2.1).

Table 8 – The Landmark Point Block

Field	Size	Value	Notes
Landmark Point Type	1 byte	See Clause 5.6.1	Denotes the type of the Landmark Point.
Landmark Point Code	1 byte	See Clause 5.6.2	Denotes the Landmark Point, e.g. the left eye.
X coordinate, Y coordinate	2 bytes	See Clause 5.6.1, Table 9	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 5.6.1, Table 9	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.

5.6.1 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 the 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 for future definition of landmark point types.

Table 9 — 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	0x04-0xFF	Reserved for future use.

5.6.2 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 5.6.3, 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 5.6.4 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 5.6.5 shall be stored in this block.

5.6.3 MPEG4 Feature 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*.

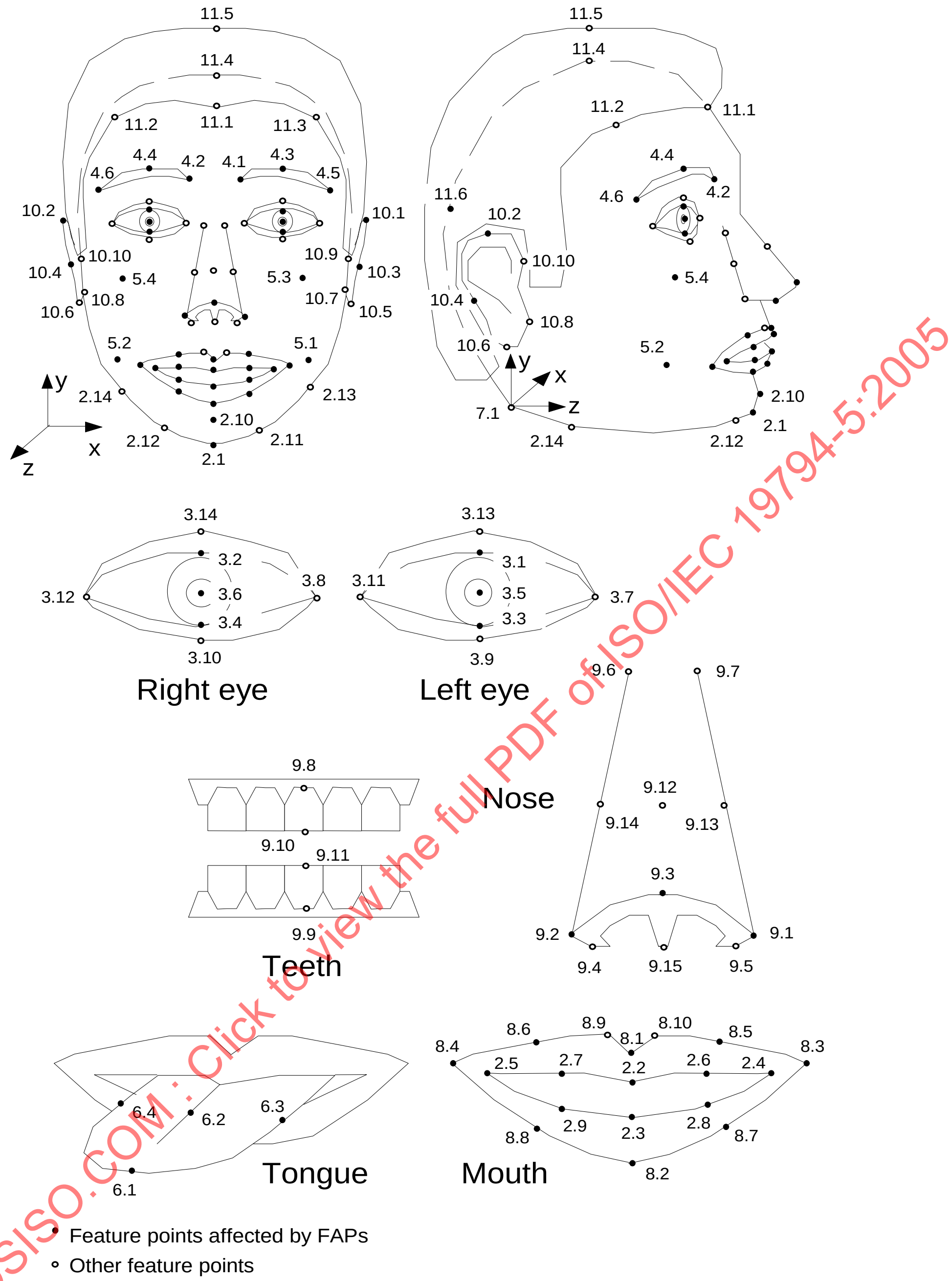


Figure 7 – The MPEG4 Feature Point Codes defined in ISO/IEC 14496-2.

Each *Landmark Point Code* in Figure 8 is given by major value *A* and minor value *B*. For example, the code for the left corner of the left eye is given by major value 3 and minor value 7.

5.6.4 Eye and nostril 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 10.

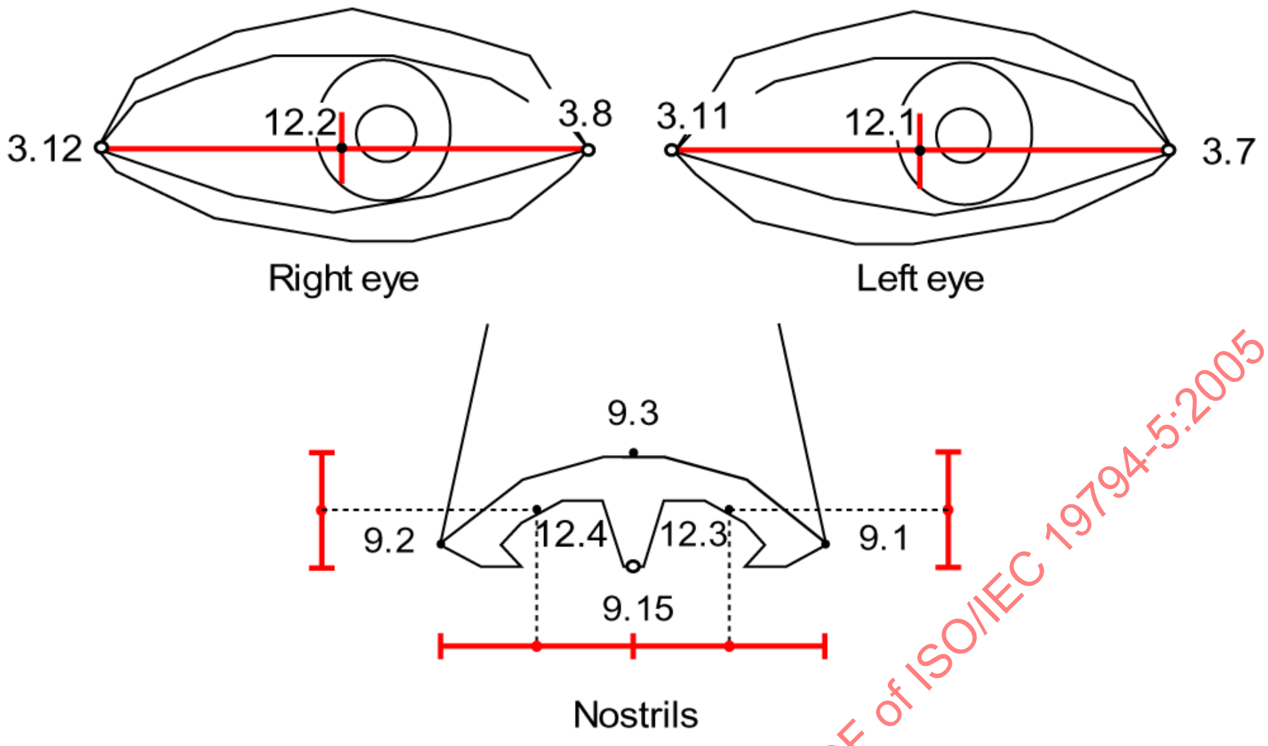


Figure 8 – The eye and nostril centre landmark points are defined by midpoints of MPEG4 feature points.

Table 10 – 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	

5.6.5 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. Figure 9 and Table 11 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. Definitions for these points are presented in the Table 11.

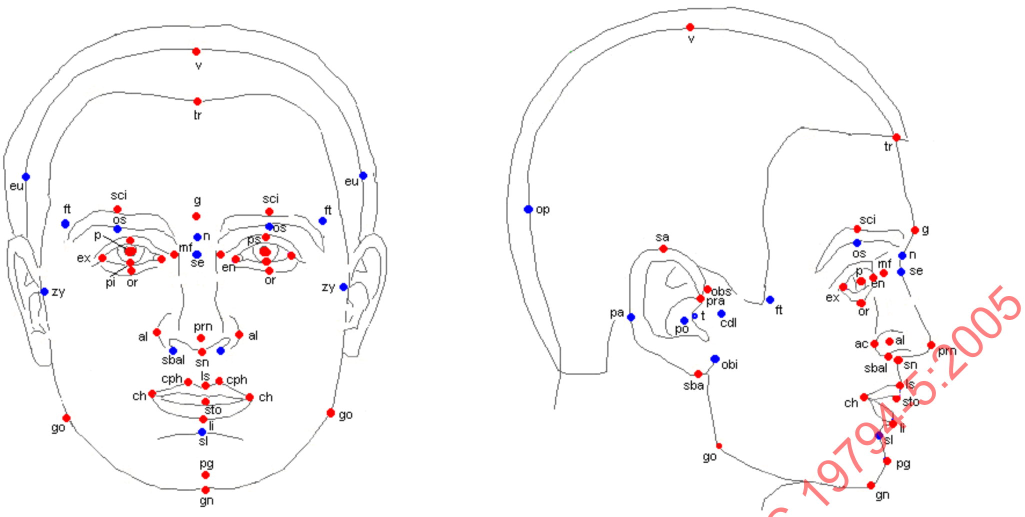


Figure 9: Anthropometric landmarks with (red) and without (blue) MPEG4 counterparts.

Table 11 – 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 C 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	The lowest median landmark on the lower border of the

			(or gnathion)	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 postrerior 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.

5.6.6 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.

5.6.7 Z Coordinate

This field is not used if the *Landmark Point Type* is equal to the MPEG4 feature or the anthropometric 2D landmark. In case the *Landmark Point Type* is an 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: 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

The (12 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 *Image Data Block* shall follow this block.

5.7.1 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 12.

NOTE: All *Frontal Image Types* are either *Full Frontal*, *Token Frontal*, or one of the respective *3D Full Frontal* or *3D Token Frontal Image Types*. Therefore a separate *Frontal* value is not required.

Table 12 – Face Image Type codes

Description	Value
Basic	0x00
Full Frontal	0x01
Token Frontal	0x02
Reserved	0x03 – 0x7F
Basic 3D	0x80
Full Frontal 3D	0x81
Token Frontal 3D	0x82
Reserved	0x83 - 0xFF

The *Basic Face Image Type* is defined in Clause 6. The *Frontal*, *Frontal/Full* and *Frontal/Token Face Image Types* are defined in Clauses 7, 8, and 9 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. Relations are indicated by an arrow from the child to the parents. 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. The 3D image types are defined in Clauses 10, 11, and 12 respectively.

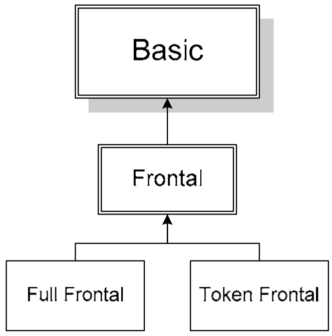


Figure 10 – 2D Face Image Types and their inheritance map. Normative requirements for the Basic, Frontal, Full Frontal and Token Frontal Face Images Types are given in Clauses 6, 7, and 8, respectively.

5.7.2 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) or JPEG2000 (ISO/IEC 15444-1) shall be specified.

NOTE: The value *Unspecified* cannot be encoded.

Table 13 – Image Data Type codes

Description	Value
JPEG	0x00
JPEG2000	0x01
Reserved	0x02 – 0xFF

5.7.3 Width

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

5.7.4 Height

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

5.7.5 Image Colour Space

The (1 byte) *Image Colour Space* field indicates the colour space used in the encoded *Image Data Block* according to the values in Table 14. The values of 128-255 are vendor specific. Application developers may obtain the values for these codes from the vendor.

Table 14 – Colour Space codes

Colour Space	Value
Unspecified	0x00
24 bit RGB	0x01
YUV422	0x02
8 bit greyscale	0x03
Other	0x04
Reserved	0x05 – 0x7F
Vendor Specific	0x80 – 0xFF

5.7.6 Source Type

The (1 byte) *Source Type* field denotes the classification of the source of the captured image and is given in Table 15. The values of 128 - 255 are vendor specific. Application developers may obtain the values for these codes from the vendor.

Table 15 - Source Type codes

Description	Value
Unspecified	0x00
Static photograph from an unknown source	0x01
Static photograph from a digital still-image camera	0x02
Static photograph from a scanner	0x03
Single video frame from an unknown source	0x04
Single video frame from an analogue video camera	0x05
Single video frame from a digital video camera	0x06
Unknown	0x07
Reserved	0x08 – 0x7F
Vendor Specific	0x80 – 0xFF

5.7.7 Device Type

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

5.7.8 Quality

The (2 byte) *Quality* field shall be reserved for future definition to represent a quality of the facial image. This field shall be set to the value 0 indicating *unspecified*.

5.8 The Image Data Block

5.8.1 Data structure

The (variable byte) *Image Data Block* shall be the image data encoded by either the JPEG or JPEG2000 standards.

Table 16 – Image Data structure

Field	Size	Value	Notes
Image Data	Variable	See Table 13 in Clause 5.7.2	Either JPEG or JPEG2000

5.9 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*, the *3D Source Type*, 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*.

5.9.1 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.

5.9.2 Coordinate System Type

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 into metric 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 metric coordinates, one has to interpolate. The original data is not preserved. This may be sufficient for many applications, but this document is intended to be able to store the original data, too.

Thus this document 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: The coordinate system may be further restricted for different *Face Image Types* (see Clauses 10 to 12, and Annex A.7).

The transformation to metric coordinates is described by appropriate scaling factors and implicit rules (e.g. as used in the 3D anthropometric landmark type) defined in Clause 5.9.2.1 to 5.9.2.2. The (1 byte) *Coordinate System Type* field specifies the coordinate system of the 3D data by using the following values.

Table 17 – The Coordinate System Type

Description	**Value**
Cartesian Coordinate System	0x00
Cylindrical Coordinate System	0x01
Reserved	0x02 – 0xFF

The different coordinate systems are defined as follows:

5.9.2.1 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 11.

NOTE: For certain *Image Types* the pose of the head is restricted. The frontal pose is defined by the Frankfurt Horizon FH (see Annex C) 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.

NOTE: 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*.

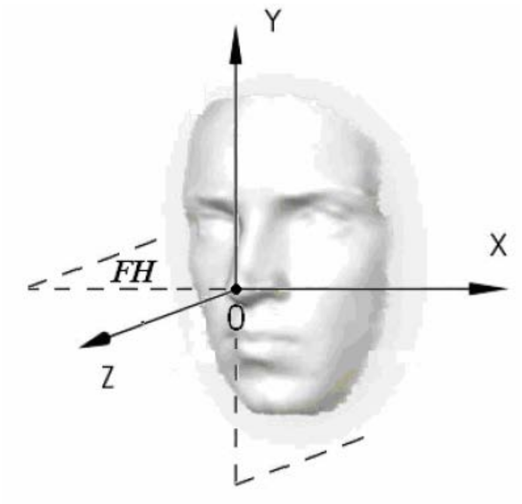


Figure 11 – A sample of a Cartesian coordinate system

5.9.2.2 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 millimeter. Note, that large values of *ScaleX, ScaleY* or *ScaleZ* indicate a low resolution 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.

For certain *Image Types* the origin of the cylindrical coordinate system shall be the nose, i.e. the *prn* (pronasale) landmark as defined in Table 11.

NOTE: 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 5.9.2.2 and then inverting the transformation given in 5.9.2.1.

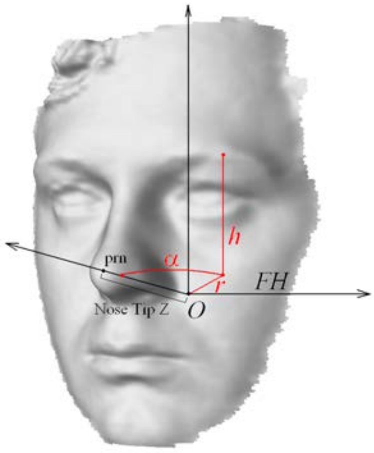


Figure 12 – Sample of a Cylindrical coordinate system. FH is the Frankfurt Horizon (Annex C).

5.9.3 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: The obtained coordinates are floating point values. In this document 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.

5.9.4 ScaleX, ScaleY, ScaleZ, OffsetX, OffsetY, OffsetZ

As outlined in Clause 5.9.2.1 and 5.9.2.2 *ScaleX, ScaleY, ScaleZ, OffsetX, OffsetY* and *OffsetZ* are needed to transform digital coordinates to metric coordinates. This applies to all three 3D representations. The values are given in millimeters. In the case of Cartesian coordinates *ScaleX* also has the unit millimeter, 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.

NOTE: Large values of *ScaleX, ScaleY* or *ScaleZ* indicate a low resolution in the respective dimension.

NOTE: Boundary values of *ScaleX, ScaleY* and *ScaleZ* may be strongly restricted for different *Face Image Types* (see Clauses 10 to 12, and Annex A.7).

NOTE: *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.

5.9.5 3D Representation Type

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

Table 18 – 3D Representation Type

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

5.9.6 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 19 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 19 – 3D Supplemental Data

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

5.9.7 3D Source Type

In analogy to the *Source Type* field in the *2D Image Information Block*, where the source of the 2D data can be coded, the (1 byte) *3D Source Type* 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 20 – The 3D Source Type

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	0x87-0xFF	0x07 – 0x80

5.9.8 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* ID is unspecified. Application developers may obtain the values for these codes from the vendor.

5.9.9 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 *2D Image Texture* 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. 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 is coded in the two's complement system. So, a value of 0x8001 codes the maximum negative temporal difference of −32767 ms and the value *0x7FFF* corresponds to the maximum positive temporal difference of +32767 ms. A value of *0x8000* is acceptable and indicates that the *3D Temporal Synchronicity* is unspecified.

Table 21 – The 3D to 2D Image Temporal Synchronicity

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

5.9.10 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 *2D Texture Map* acquisition and the start of the 3D acquisition process in milliseconds.

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 is coded in the two's complement system. So, a value of *0x8001* codes the maximum negative temporal difference of −32767 ms and the value 0x7FFF corresponds to the maximum positive temporal difference of +32767 ms. A value of 0x8000 is acceptable and indicates that the *3D to 2D Texture Temporal Synchronicity* is unspecified.

Table 22 – 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 in two's complement coding.	0x0000 – 0x7FFF 0x8001 – 0xFFFF
Unspecified	0x8000

5.9.11 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 milliseconds. A value of *0xFFFF* is acceptable and indicates that the field is not specified.

Table 23 – The 3D Acquisition Time

Description	**Value**
Duration of the 3D acquisition process in milliseconds.	0x0000 – 0xFFFE
Unspecified	0xFFFF

5.9.12 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 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 24 – The 2D Texture Acquisition Time

Description	**Value**
Duration of the 2D acquisition process in milliseconds.	0x0000 – 0xFFFE
Unspecified	0xFFFF

5.9.13 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 25 – The Texture Map Type codes

Description	**Value**
Unspecified	0x00
JPEG	0x01
JPEG2000	0x02
PNG	0x03
Reserved	0x04-0xFF

5.9.14 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 5.10.8. 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 Data 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 T*exture Map Spectrum* field shall be *unspecified*.

Table 26 – 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	0x05-0xFF

5.10 The 3D Data Block

The *3D Data Block* contains the representation of the 3D data. There are three alternatives to store 3D Data: range image, 3D point map, and vertex data. Optionally, additional information can be stored in the *Image Mask* 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 5.9.5) is used to define the format of the 3D data representation that has been used in the actual record.

5.10.1 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 27 – The Range Image Bit Depth codes

Description	**Value**

8 bit	0x00
16 bit	0x01
Reserved	0x02-0xFF

5.10.2 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 image 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 5.10.1). 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.

5.10.3 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 65535.

5.10.4 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.

5.10.5 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 10.3.2.

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 29 in Clause 5.10.7. 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 28 – The Normal Flag codes

Description	Value
Normal information not used in Vertex Data	0x00

Normal information used in Vertex Data	0x01
Reserved	0x02-0xFF

5.10.6 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 5.10.5). 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.

5.10.7 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 200 are reserved for future use. Values of $t = 201$ and above code a specific potential or corrected defect of the 3D data or the corresponding *Texture Image*.

See Clause 5.10.5 of how the pixel values are used in for the 3D vertex representation.

Table 29 – The Error Map

Description	**Value**
Reserved for future use	0…200
Depth value is interpolated, interpolation type is not specified	201
Depth value is interpolated, linear interpolation has been used	202
Depth value is interpolated, bi-cubic 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	206…255

5.10.8 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 2D facial record data 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*. Conformance requires that the 2D images of the image data block of any 3D image type have to be conformant with the relevant specifications as specified in Clauses 5 to Clause 9. See Clause 10, 11 and 12 for more details.

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 two possible encodings is to be used for all image types: The JPEG Sequential baseline (ISO/IEC 10918-1) mode of operation and encoded in the JFIF file format (the JPEG file format) or the JPEG-2000 Part-1 Code Stream Format (ISO/IEC 15444-1) and encoded in the JP2 file format (the JPEG2000 file format).

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 Faces* fields shall be specified.

6.4.2 Facial Information

The *Block Length* and *Number of Feature 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.

NOTE: 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* are either *Full Frontal* or *Token Frontal*. (Refer 5.7.1)

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 the 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° 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° from frontal in roll (ref. 5.5.8). Figure 13 shows an example of ±8° rotation in roll.

Figure 13 – Sample images with +8° (left) and -8° (right) rotation in roll.

NOTE: The best practice recommendation as outlined in Clause A.2.2 is that the rotation of the head should be less than ±5° from frontal in roll.

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.

- Neutral (non-smiling) with both eyes open normally (i.e. not wide-open), and mouth closed.
- A smile where the inside of the mouth and/or teeth is not exposed (closed jaw).
- A smile where the inside of the mouth and/or teeth is exposed.
- Raised eyebrows.
- Eyes looking away from the camera.
- Squinting.
- Frowning.

See the 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 the 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 the 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 7.2.8 and 7.2.9

7.2.8 Shadows over the face

The region of the face, from the crown (as defined in Clause 4.6) to the base of the chin, and from ear to ear, shall be clearly visible and free of shadows. Special care shall be taken in cases when veils, scarves or headdresses cannot be removed for religious reasons to ensure these coverings do not obscure any facial features and do not generate shadow. In all other cases head coverings shall be absent.

7.2.9 Shadows in eye sockets

There shall be no dark shadows in the eye sockets due to the brow. The iris and pupil of the eyes shall be clearly visible.

7.2.10 Hot spots

Care shall be taken to avoid hot spots (bright areas of light shining on the face). These artefacts are typically caused when one, high intensity, focused light source is used for illumination. Instead, diffused lighting, multiple balanced sources or other lighting methods shall be used. A single bare point light source is not acceptable for imaging. Instead, the illumination should be accomplished using other methods that meet requirements specified in this Clause.

7.2.11 Eye glasses

If the person normally wears glasses then they should wear glasses when their photograph is taken. Glasses shall be clear glass and transparent such that the eye pupils and irises are clearly visible. 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. Permanently tinted glasses or sunglasses are acceptable only for medical reasons (and shall otherwise be removed). In cases where tinted glasses or sunglasses are worn, the specification of dark glasses in the header structure is recommended. Care shall be taken that the glasses frames do not obscure the eyes. There shall be no lighting artefacts or flash reflections on glasses. This can typically be achieved by increasing the angle between the lighting, subject and camera to 45º (degrees) or more.

7.2.12 Eye patches

The wearing of eye patches is allowed only for medical reasons. In these cases, the specification of the patch, in the header structure is recommended.

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 the Informative Annex A.2 *Best Practices for Frontal Images*.

7.3.2 No over or under exposure

For each patch of skin on the person's face, the gradations in textures shall be clearly visible. In this sense, there will be no saturation (over- or underexposure) 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. 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 will have sufficient depth of focus to maintain greater than 2 mm resolution on the subject's facial features at time of capture.

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. Red-eyes are not acceptable.

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. 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 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.

7.4.2.2 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.

7.4.2.3 Colour space

Frontal images shall be represented as one of the following

- 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.

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

- 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.

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, 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).

7.5 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 and Clause 7.

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 and Clause 7.

8.3 Photographic requirements for the Full Frontal Face Image Type

8.3.1 Introduction

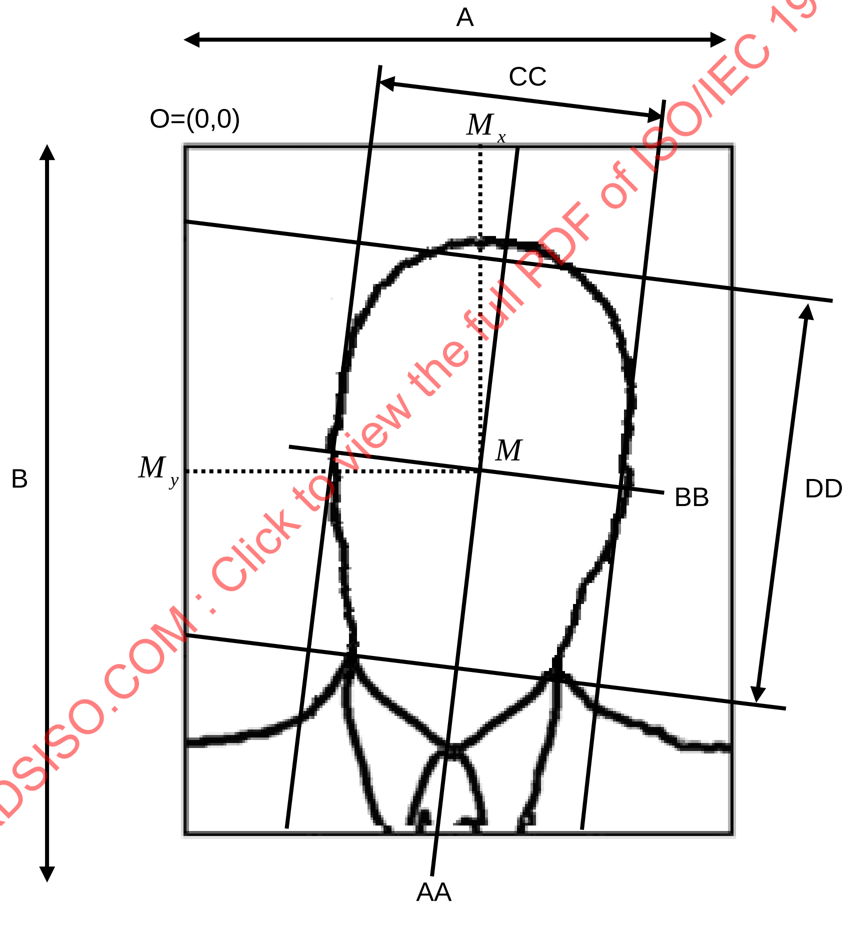


Figure 14 – Geometric characteristics of the Full Frontal Face image. This figure is a derivative of AAMVA document DL/ID-2000.

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

NOTE: 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 resolution and the photographic requirements of this Clause.

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). The origin *O* of the coordinate system is defined to be 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. 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*).

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 30 summarizes the photographic requirements for full frontal images specified in Clauses 8.3.1 – 8.3.5.

Table 30 — 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	$0.45\ A \le M_x \le 0.55\ A$
8.3.3	Vertical Position of Face	$0.3\ B \le M_y \le 0.5\ B$
8.3.3	Vertical Position of Face (Children under the age of 11)	$0.3\ B \le M_y \le 0.6\ B$
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
8.3.5	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.

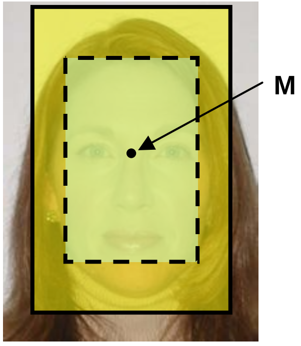


Figure 15 — Sample image with the respective minimal and maximal head dimensions based on the image width and height.

8.4 Digital requirements for the Full Frontal Face Image Type

8.4.1 Resolution

For an image for optimal human examination and permanent storage, the resolution of the full images shall be at least 180 pixels of resolution for the width of the head, or roughly 90 pixels from eye centre to eye centre. See the Informative Annex Clause A.3.1.1 for best practices on this topic.

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 *0x01*.

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 and Clause 7.

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 *Feature Points* 12.1 and 12.2, shall be determined. For the determination of eye positions, it is possible:

- To use computer inspection,
- To use human visual inspection, or
- To use computer and human visual inspection.

9.2.3 Token image geometric format

A *Token Image* is a colour or greyscale image with image dimensions and eye position coordinates given by Table 31.

NOTE: Clause 5.2.3 specifies conversion of values to integer.

Table 31 – 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.

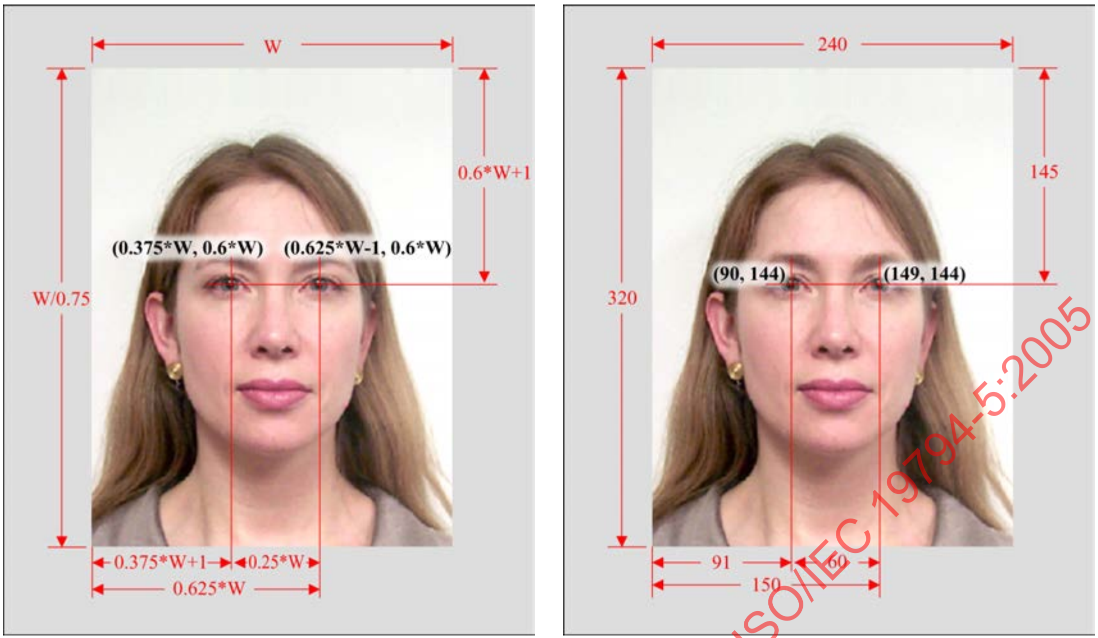


Figure 16 – The Token Face Image Type geometric format (left) and a sample minimum width (*W*=240) Token Face Image format image (right).

9.2.4 Minimum width Token image

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 the A.4.3 for best practices on this matter.

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 Basic 3D Image Type

10.1 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 Clause 10. 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: 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.

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

10.2.1 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.

10.2.2 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

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

10.3.1 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.

10.3.2 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

11 The Full Frontal 3D Image Type

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

11.1 Inheritance requirements

The *Full Frontal with 3D Image Type* inherits the requirements of the *Basic 3D Image Type* as specified in Clause 10.1. Furthermore, it inherits all requirements of the *Full Frontal Image Type*.

11.2 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 11.

11.3 Pose of the 3D representation

The rotation of the head in the 3D representation shall be less than ±5° from frontal in pitch and yaw (see Clause 5.5.6 and 5.5.7). 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° from frontal in roll (ref. Clause 5.5.8). 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.

11.4 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 5.9.9) will be indicative of the observed effect.

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

11.5.1 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: In any case, *ScaleX* and *ScaleY* denote a sampling rate, not the physical measurement rate of the sensor.

11.5.2 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.

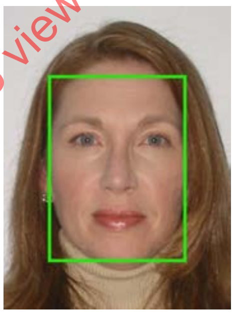


Figure 17 – Sample 2D image of the minimal Face Coverage (outer region).

11.5.3 Non-valid points in 3D data Image

At maximum 50% of the pixels of the *Range Image* in the region defined in 11.5 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 pixels shall have a zero value, indicating an invalid 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.

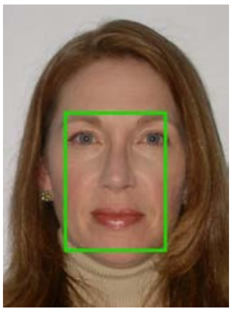


Figure 18 – Sample 2D image of the inner region.

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

11.6.1 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.

11.6.2 Face Coverage

For biometric purposes it is eminently important that the *outer region*, as defined in Clause 11.5, 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 $-1.75\ w \le X \le 1.75\ w$ and $-1.75\ w \le Y \le 2.55\ w$ 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 5.6.4.

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

11.7.1 Face Coverage

For biometric purposes it is eminently important that the inner region, as defined in Clause 11.5, 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 $-1.5\ w \le X \le 1.5\ w$ and $-1.8\ w \le Y \le 1.8\ w$ 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.

12 The Token Frontal 3D Image Type

12.1 General

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

12.2 Inheritance requirements

The *Token Frontal 3D Image Type* inherits all requirements of the *Basic Image Type*, the requirements of the *Full Frontal Image Type* as defined in Clauses 11.2 to 11.6. Furthermore, it inherits all requirements of the *Token Frontal Image Type*.

12.3 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 11.7.

12.4 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 11.8.

12.5 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 11.9.

Bibliography

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

[2] ISO/IEC 19784-1:2006 – Information Technology – Biometric application programming interface – Part 1: 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] 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

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

[8] [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).

[9] 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

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

[15] ISO 10526:1999/CIE S005/E-1998, CIE Standard Illuminants for Colorimetry

Annex A

(informative)

Best practices

A.1 Best practices for 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.

A.1.2 Feature Point determination

The *Feature 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 feature points (landmarks) used by face recognition algorithms. If possible, feature points should be determined on images before compression is applied.

Feature points should be included in the record format if they have been accurately determined, thereby providing the option that 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 feature 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 feature point determination followed by human verification and potential override of the computer determined feature points.

A.2 Best practices for 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.

A.2.2 Pose

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

A.2.3 Expression

The expression should be neutral (non-smiling) with both eyes open normally (i.e. not wide-open), and mouth closed. Every effort should be made to have supplied images comply with this specification. A smile with closed jaw is not recommended. Examples of unacceptable expressions are closed eyes, hair covering eyes, and rim of glasses covering part of the eye. Examples of non-recommended expressions are a smile where the inside of the mouth and/or teeth is exposed (jaw open), raised eyebrows, eyes looking away from the camera, squinting, and frowning.

A.2.4 Assistance in positioning the face

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

A.2.5 Background

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.2.5.1 Background segmentation

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

A.2.5.2 Background shadows

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

A.2.5.3 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.2.5.4 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.2.6 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 resolution 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.2.7 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.2.8 Colour calibration

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

A.2.9 Radial 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.

A.3 Best practices for Full Frontal Images

A.3.1 Digital attributes of Full Frontal Images

A.3.1.1 Photo resolution

For an image for optimal human examination and permanent storage, the preferred minimum resolution of the full image is at least 240 pixels of resolution 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.

For a photograph with head width 20 mm (roughly 0.78 inches), the recommended scanner resolution is 120 dots per centimetre (roughly 300 dots per inch). For a photograph with head width 13 mm (roughly 0.5 inches), the recommended scanner resolution is 189 dots per centimetre (roughly 480 dots per inch). 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 resolution 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 resolution should typically be set to 300 dpi.

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

A.3.2 Best practices for use of Full Frontal Images on Travel Documents

A.3.2.1 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 drivers' license images.

A.3.2.2 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.3.2.3 Summary of best practice photographic recommendations

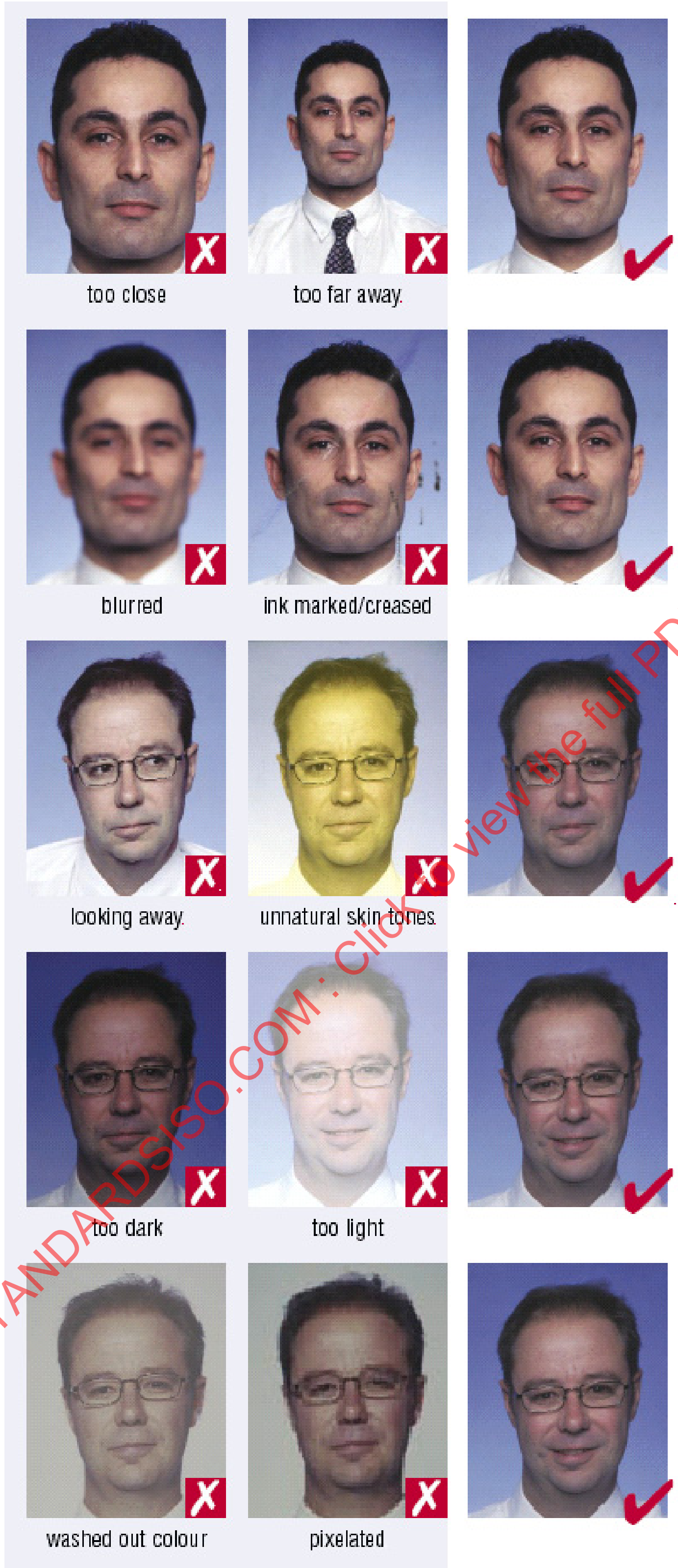
For convenience, Table A.1 summarizes the geometric and pose constraints in Clauses A.2.2 and A.3.2.1-A.3.2.2.

Table A.1: Summary best practices for Full Frontal Images on Travel Documents

Clause	Definition	Recommendation
A.2.2	Pose	±5° from frontal in roll, pitch and yaw
A.3.2.1	Width to Height Ratio of Image	$1.25 \le B/A \le 1.34$

A.3.2.2	Width of Head	1.4 $CC \le A \le 2$ CC
A.3.2.2	Length of Head	0.7 $B \le DD \le 0.8$ B
A.3.2.2	Length of Head (Children under the age of 11)	0.5 $B \le DD \le 0.8$ B

A.3.2.4 Sample images and sample photograph taking guidelines for travel documents



Photograph quality

The photographs must be:

- no more than 6-months old
- 35–40mm in width
- close up of your head and top of your shoulders so that your face takes up 70–80% of the photograph
- in sharp focus and clear
- of high quality with no ink marks or creases

The photographs must:

- show you looking directly at the camera
- show your skin tones naturally
- have appropriate brightness and contrast
- be printed on high quality paper, and at high resolution

Photographs taken with a digital camera must be high quality colour and printed on photo-quality paper.

Figure A.1 – Best practices for the purpose of travel document creation.

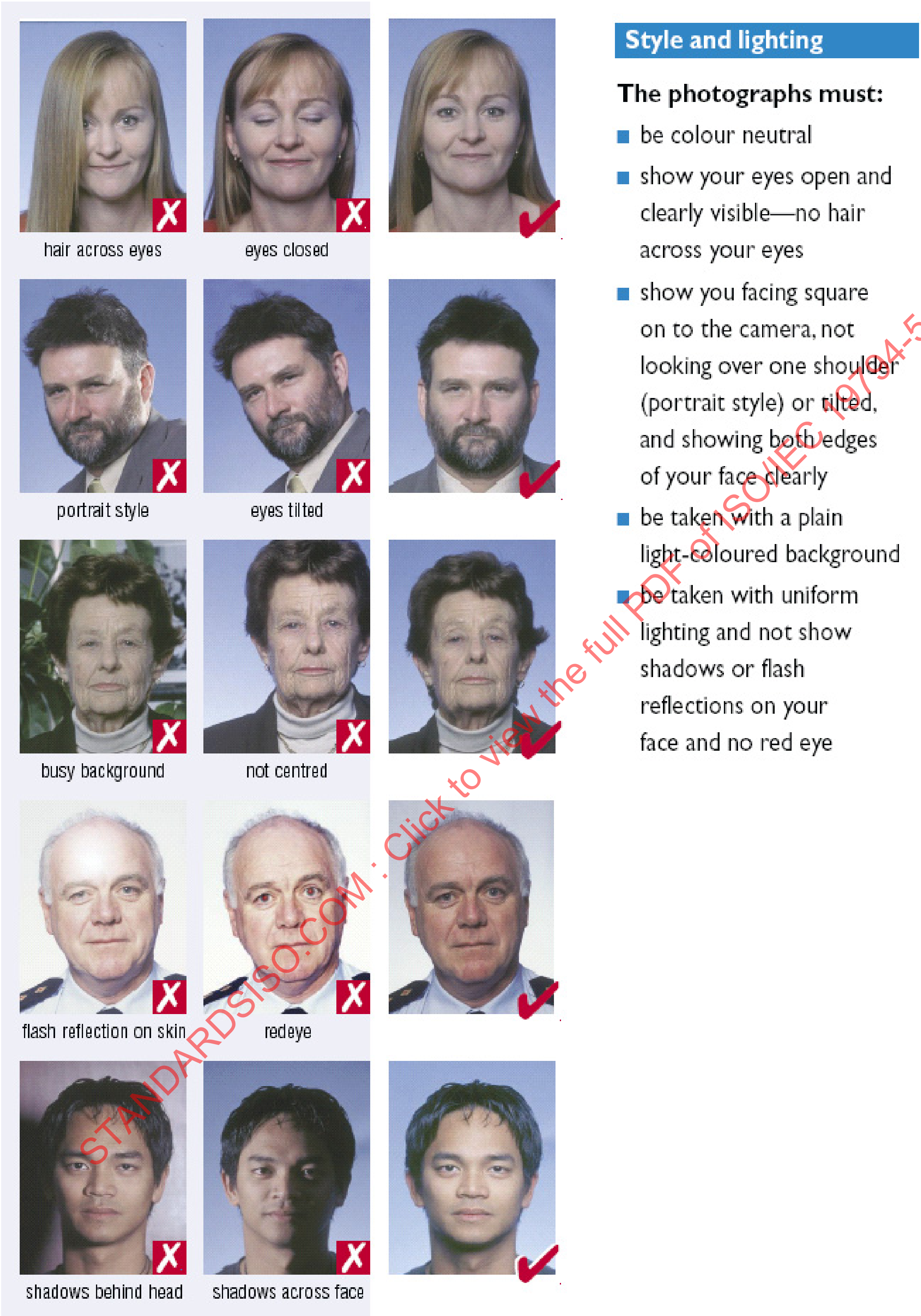


Figure A.1 – Best practices for the purpose of travel document creation.

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 A.1 – 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 Figures A.2 and 13 below, from faces obtained from Passports Australia within the Australian Department of Foreign Affairs and Trade[1]. Here, 1000 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 9. 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 identification.

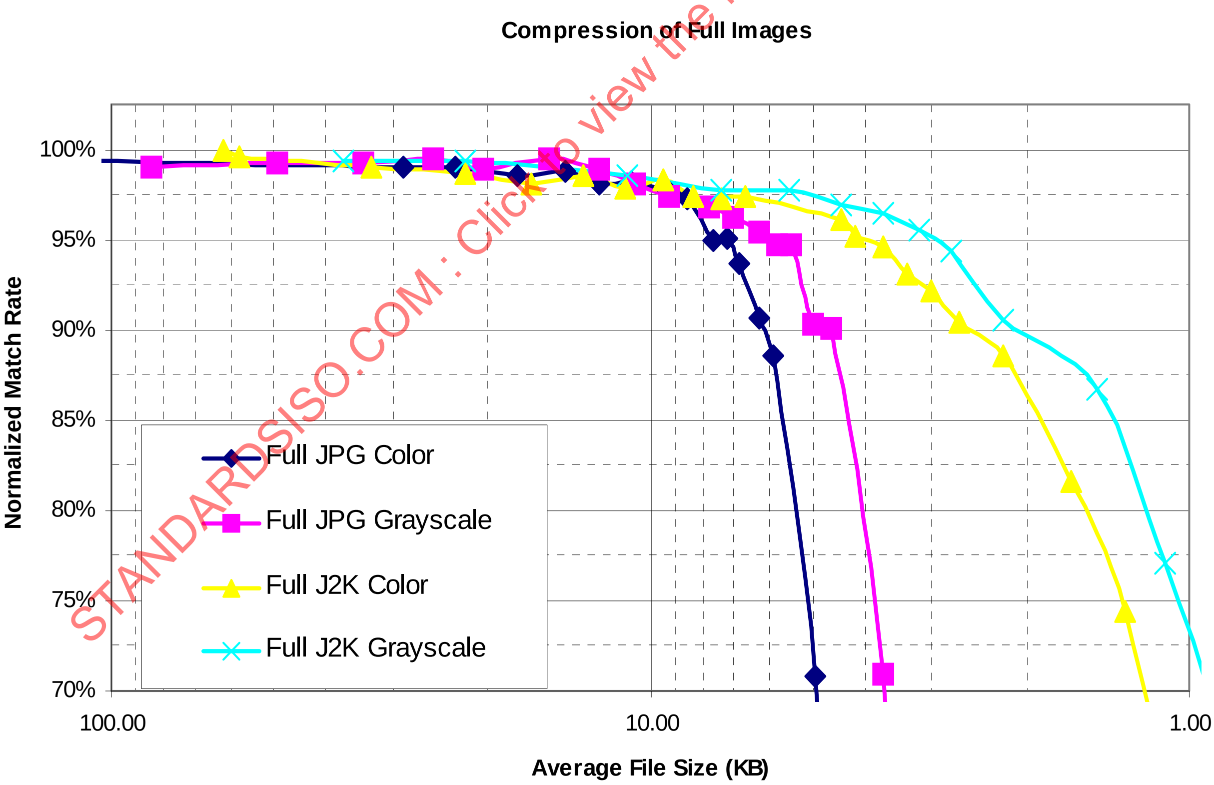


Figure A.2 – Identix, compression of Full Frontal Images versus face recognition performance.

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

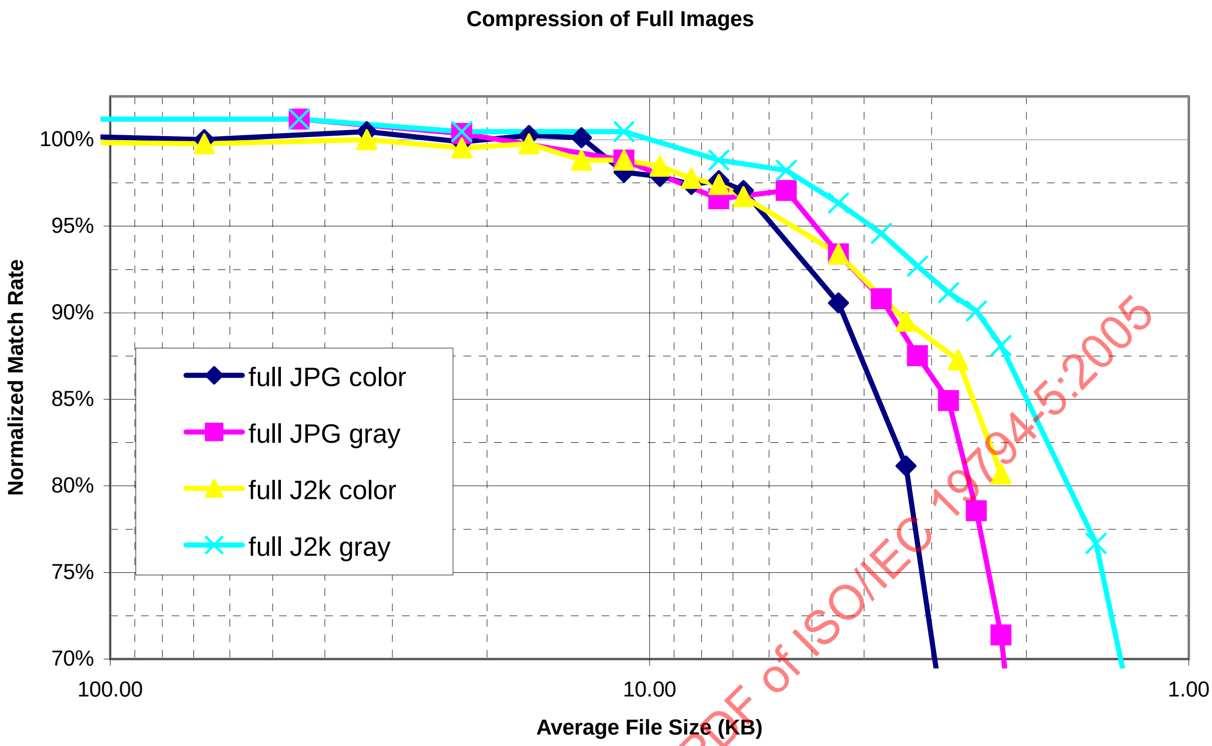


Figure A.3 – ZN Vision Technologies, compression of Full Frontal Images versus face recognition performance

In Figures A.2 and A.3, 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 1000 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 Figures A.2 and A.3 show that the performance degrades quickly below a compressed file size of 10KB, and that 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 identification (1:N searching), 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

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 A.4, a 200:1 ratio in the Outer Region leads to a (658-230.4 KB)/200 = 2.14 KB data size and a 50:1 ratio in the Inner Region to a data size of (230.4 KB)/50 = 4.61 KB, resulting in a total file size of 2.14 + 4.61 = 6.75 KB. The file size reduction is about ~40%.

In Figure A.4, 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.

Figure A.4 – Example uncompressed (left) and compressed (right) using region of interest shown left.

A.4 Best practices for 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 A.2.

Table A.2 – 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 A.5 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 with the best practices being to extend the colour used on the border of the original image to the edges of the final *Token Face Image* (e).

Figure A.5 – 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. If padding in the corners of the images is necessary, padding should be done by repetition of the last valid pixel per row for additional rows and per column for additional columns of the transformed original image. Linear interpolation between the corresponding valid horizontal and vertical pixels should be done. For regions, which cannot be filled in by the interpolation method, the original boundary values should be applied. 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.4.4 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 Figures A.6 and A.7 below, from faces obtained from Passports Australia within the Australian Department of Foreign Affairs and Trade[2]. Here, 1000 matching pairs (original and renewals) of real passport images were considered. See Clause A.2.4 for a detailed description of the compression experiments. The same experiment for *Full Frontal Images* discussed in Clause A.2.4 was repeated for *Token Images*. The results for Identix and ZN Vision Technologies are shown in Figures A.6 and A.7.

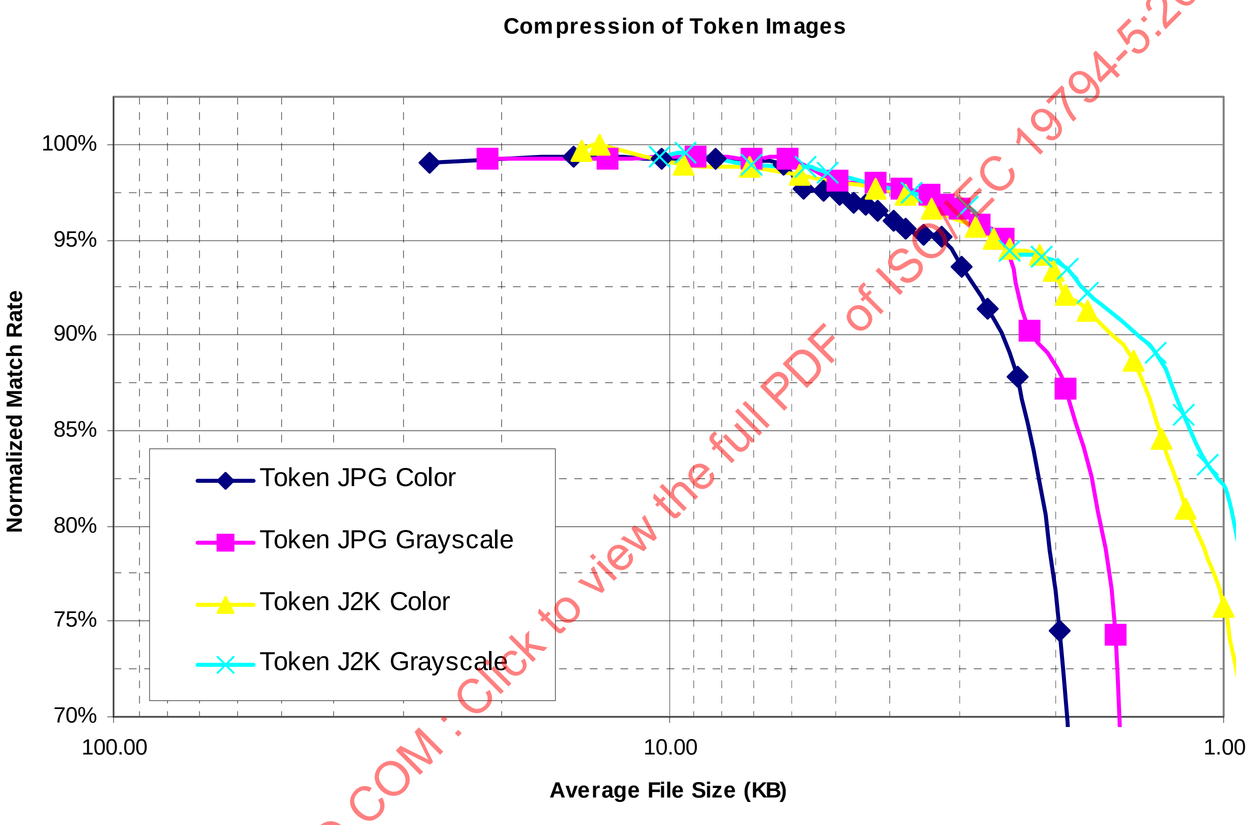


Figure A.6 – Identix, compression Token images versus face recognition performance.

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

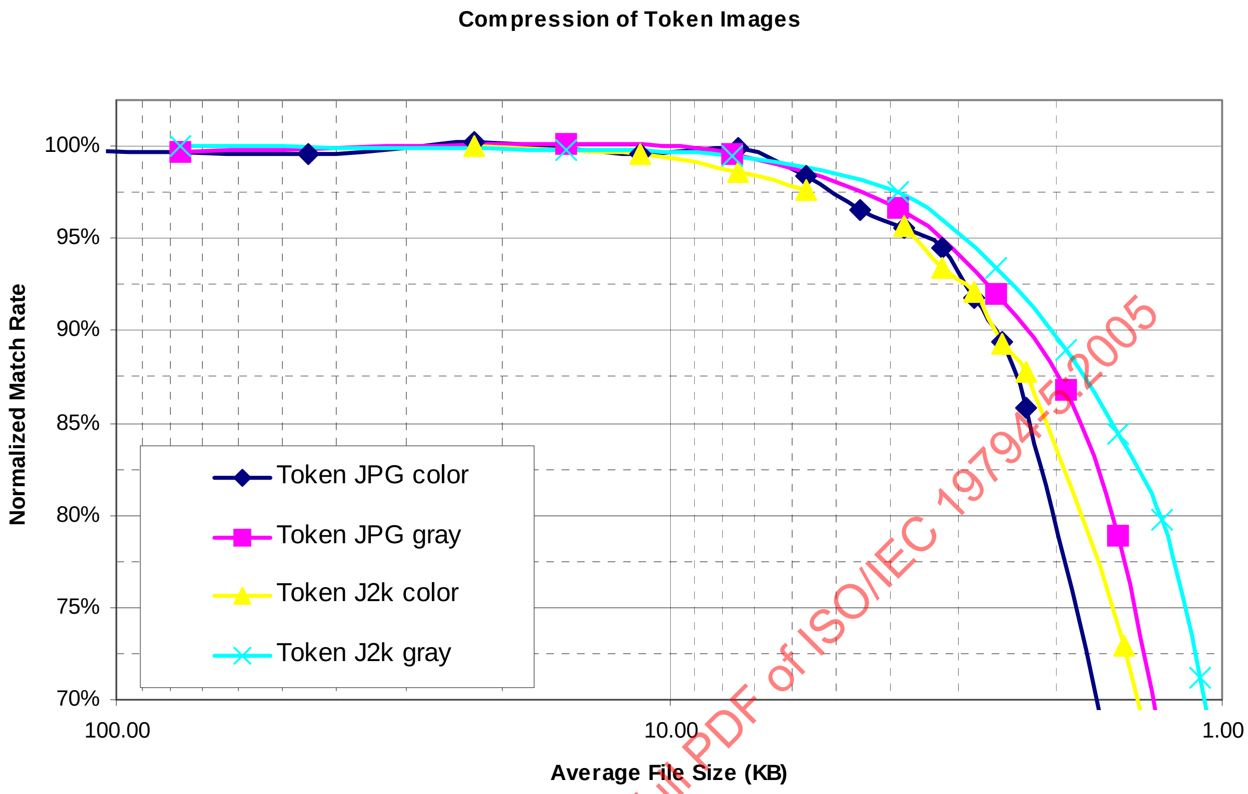


Figure A.7 – ZN Vision Technologies, compression of Token Images versus face recognition performance.

The graphs in Figures A.6 and A.7 show that the performance degrades quickly below a compressed file size of 8KB, and that 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 identification (1:N searching), 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 A.8. This is a region of 192 x 240 pixels in size. Figure A.8 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=30720 pixels, or (75-45) x 3 KB = 90 KB of size in bytes for 3 bytes per pixel.

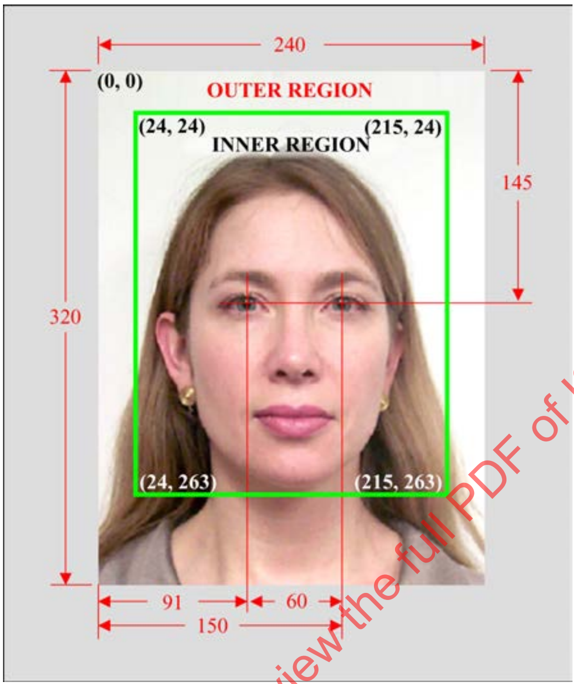


Figure A.8 – Suggested region of interest for Token Images

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

A.5 Experimental study on the enrolment of full frontal images for travel documents

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.

A.5.1 Software and data used for the analysis

The parameters and tolerances used in this study were either the strict tolerances as demanded by this document and ICAO recommendations, or the relaxed tolerances as suggested by ICAO 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 this document; in particular those with respect to pose, pixel resolution between the eye centres, relative horizontal position of the face, relative vertical position of the face, head to image width ratio and head to image height ratio.

Table A.3 — Real-world image datasets used for the analysis

Dataset Country	# of images	Pixel size (width x height)	Format
0	1000	413 x 531	JPEG
A	1988	384 x 480	JPEG
B	1911	449 x 599	JPEG
C	2229	416 x 536	JPEG

The data derived from these passport images are subsequently compared with each other and with the tolerances specified in this document.

A.5.2 Experimental results

A.5.2.1 Inter-eye distance

This document specifies a minimum of 90 pixels between eye centres for full frontal images (ref Clause 8.4.1). Figure A.9 shows the distribution of the inter-eye centres for the four sample data sets.

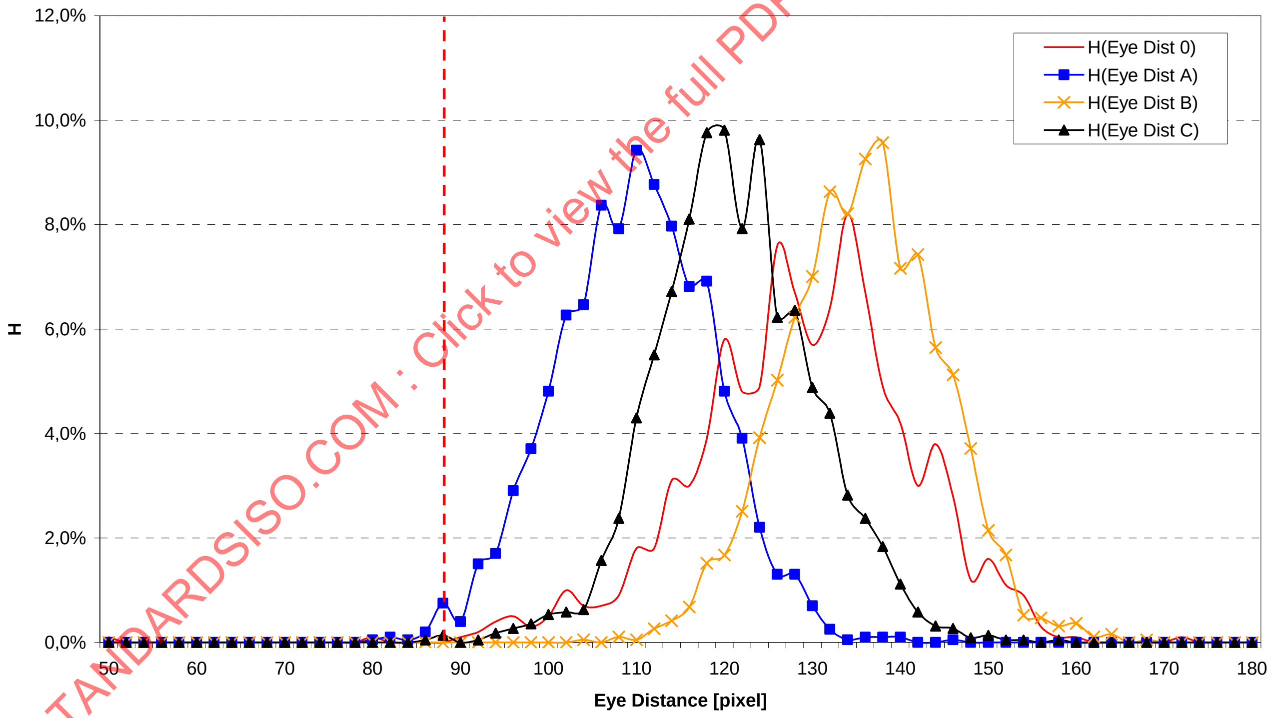


Figure A.9: Normalized distribution of the eye distance in the e-passport images of four Issuing States; the required limit is depicted

The requirement of 90 pixels between the eyes was met in almost all of the cases. The average distance between the eyes was found to be at 123.4 pixels.

A.5.2.2 Relative horizontal position of the face

This document 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 A.10 shows the distribution of the horizontal position of the face (M_x / A) for the four sample data sets.

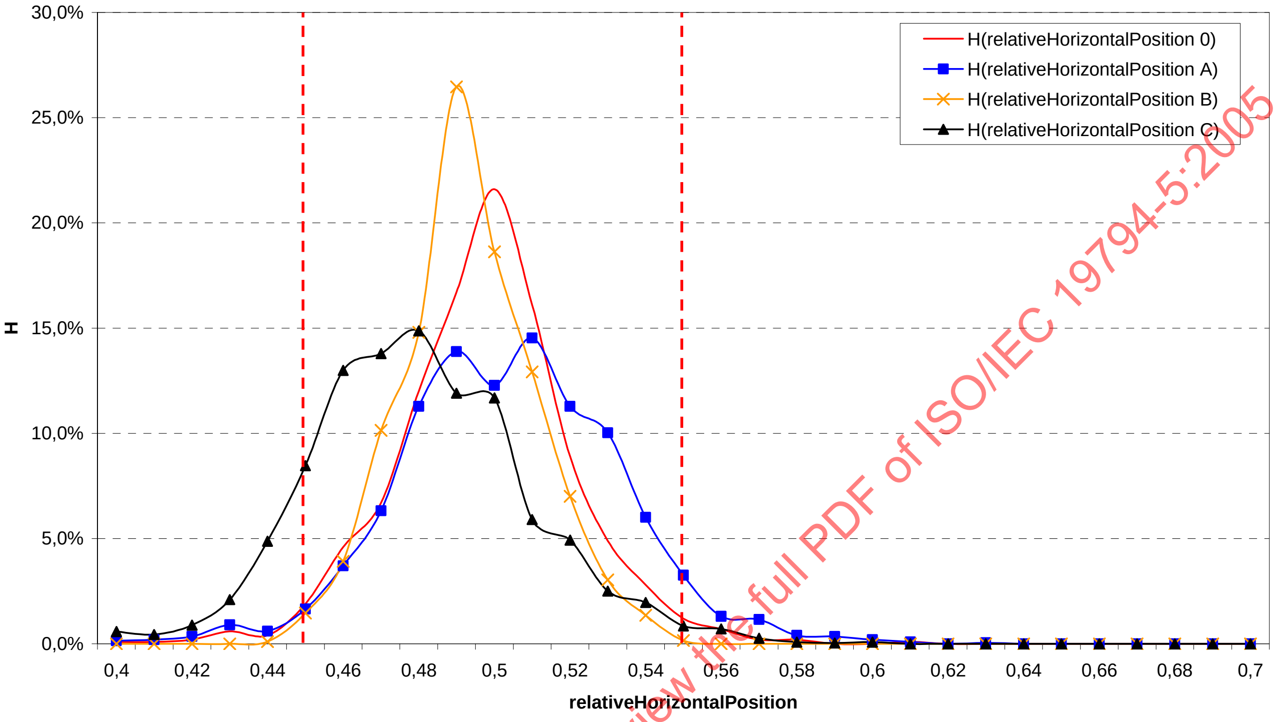


Figure A.10: Normalized distribution of the relative horizontal head position in the e-passport images of four Issuing States. The required limits are depicted.

The average horizontal head position of the 7,128 images is at approximately 49% of the image width. The specifications have been met by 95.4% of all passport photos of this study.

A.5.2.3 Relative vertical position of the face

This document 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 A.11 shows the distribution of $1 - M_y / B$ for the four sample data sets.

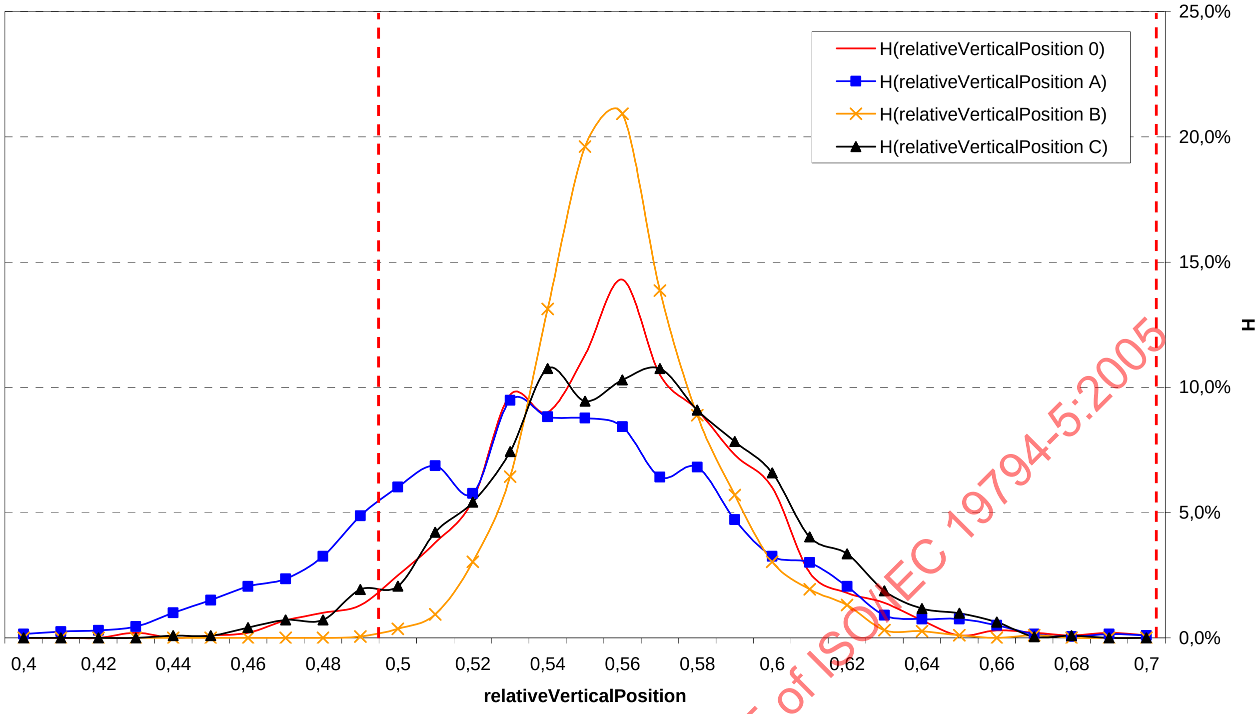


Figure A.11: Distribution of the relative vertical head position $1 - M_y / B$ in the e-passport images of four Issuing States; the limits specified are depicted

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. The original data shown in the diagram of the study recorded the distribution of ($1 - M_y / B$) instead of directly recording M_y.

A.5.2.4 Head Image Width Ratio

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 A.12 shows the distribution of the head width to image width ratio for the four sample data sets.

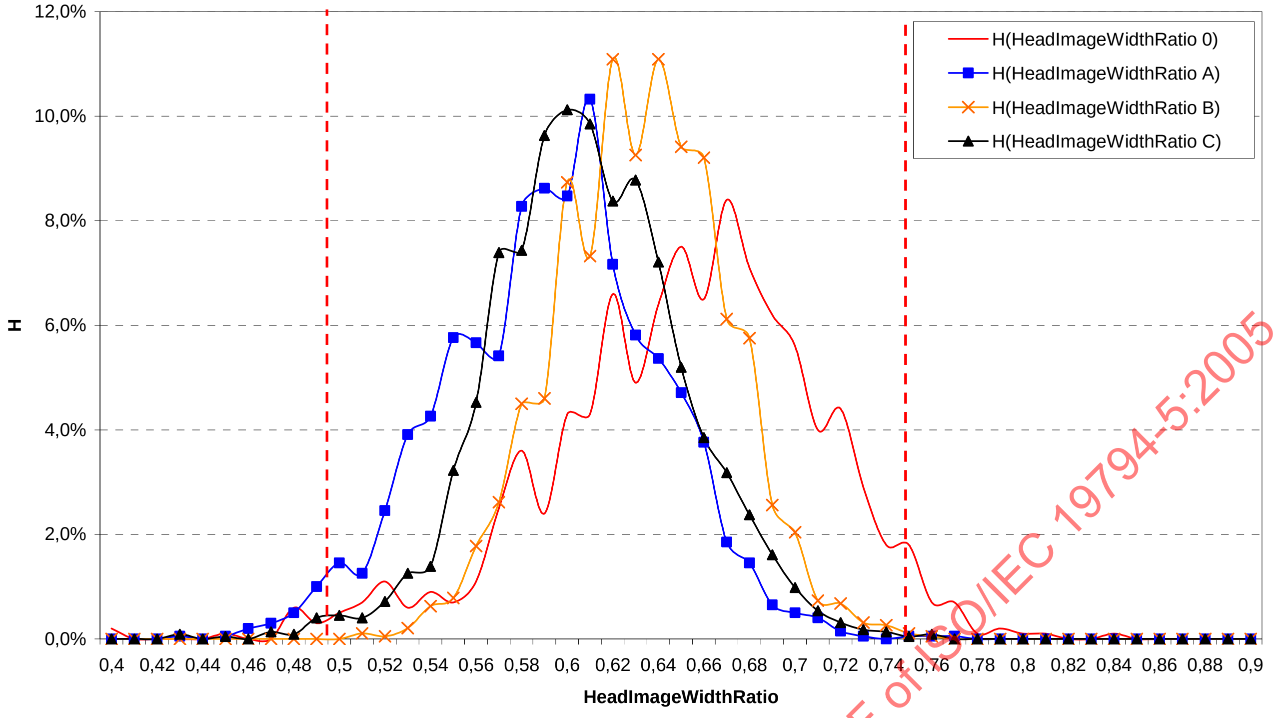


Figure A.12: Distribution of the head to image width ratio in the e-passport images of four Issuing States; the required limits are depicted.

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.

A.5.2.5 Head Image Height Ratio

In order to assure that the entire face is visible in the image it is specified 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 A.13 shows the distribution of the head height to image height ratio for the four sample data sets.

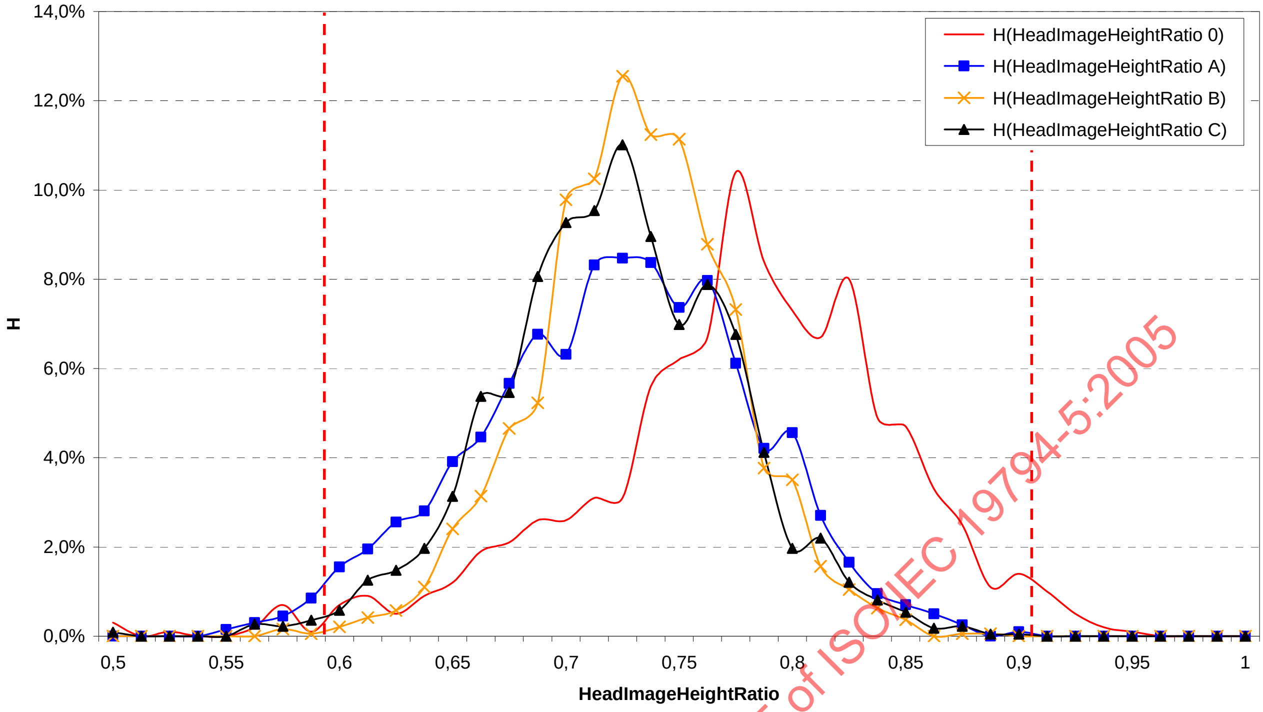


Figure A.13: Distribution of the head to image height ratio in the e-passport images of four Issuing States.

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 specified tolerances.

A.5.3 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%.

A.5.4 Summary

The study presented in this annex concentrates on geometrical parameters of the face in the photograph which are important for biometric matching 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 A.4 — Summary of the compliance of the sample to the specified requirements.

Criteria	Min	Max	Images compliant
Eyes Distance, min [pixel]	90.00		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%

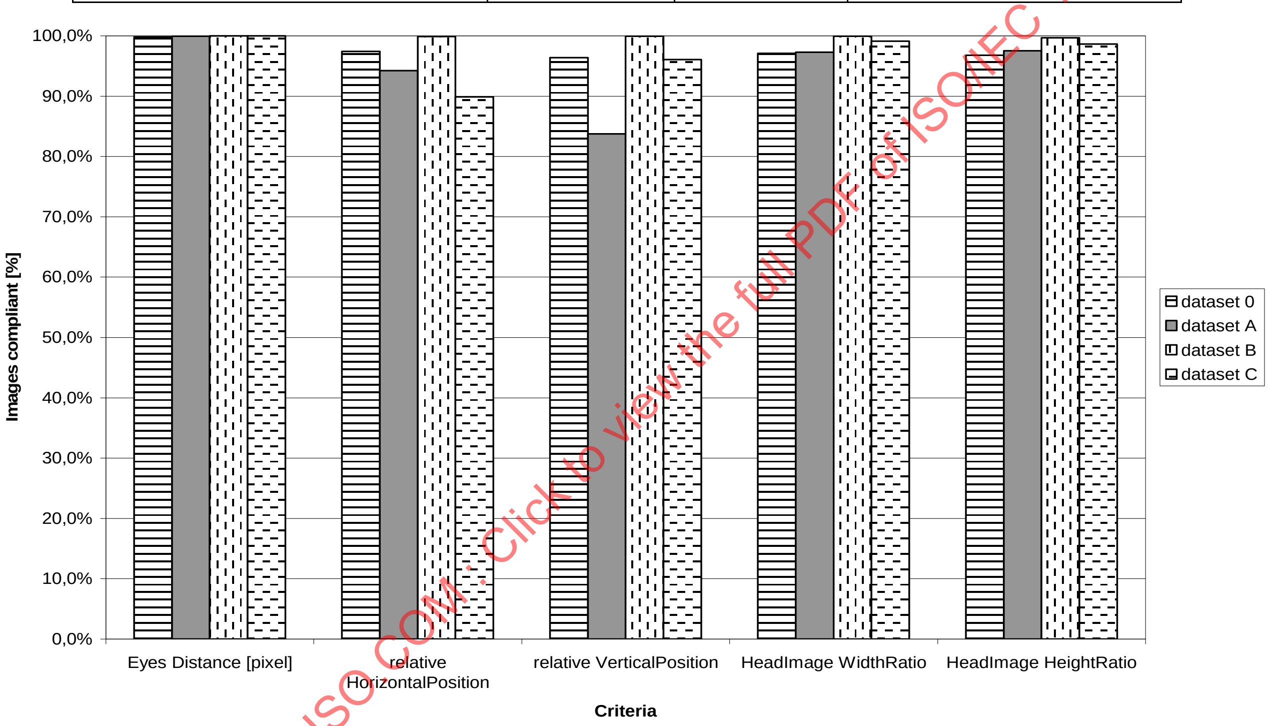


Figure A.14: Percentage of compliant passport photos in the test set of 7,128 images of four Issuing States using the tolerances as put forward in Clause 8.3.

In summary, the analysis given above shows that with current technology the thresholds can be achieved in real world applications. Table A.4 and Figure A.14 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.

A.6 Study on the effects of inter-eye distance and roll on biometric comparison performance

A.6.1 Inter-eye distance

The distance between the eyes (i.e., the resolution 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 matching algorithm. Figure A.15 shows an increasing verification rate with increasing eye distance. These results also hold for investigations of the identification performance (rank statistics).

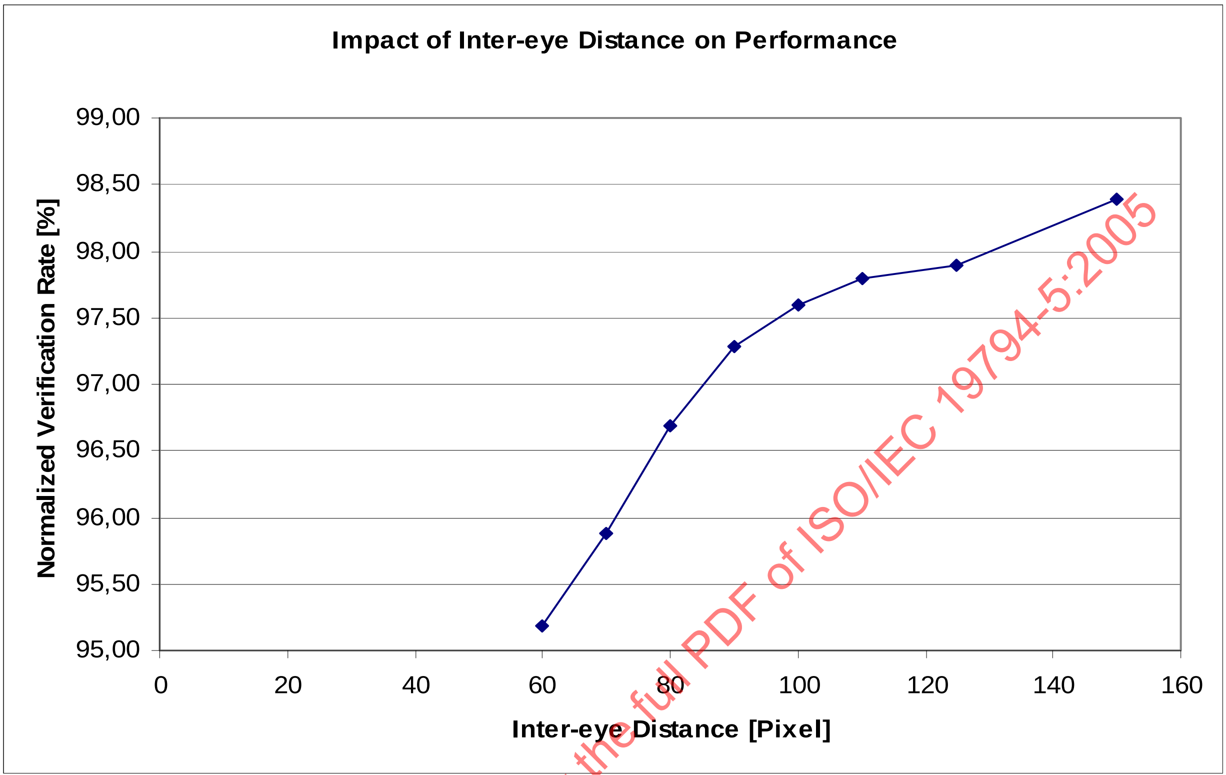


Figure A.15: Normalized verification rate vs. inter-eye distance at FAR=0.1%

A.6.2 Pose

Automated face recognition systems allow for effective compensation of the roll angle of a facial image. Figure A.16 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 angles. Then tests have been carried out, where all probes have been rotated by +5° and all gallery images by -5°, then vice versa. This has been repeated for 10°, -10°, 15° and -15°, respectively. So, the delta in roll angle of probe and gallery is 0°, 10°, 20° and 30°. Figure A.17 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: This setup is the worst case scenario as all images in the ±10° tests show more than the maximum tolerance of ±8°. 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°. These results also hold for investigations of the identification 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° roll for a single image, i.e. a maximum of 20° difference in the roll angles. This is the reason for the decrease for the ±15° tests.

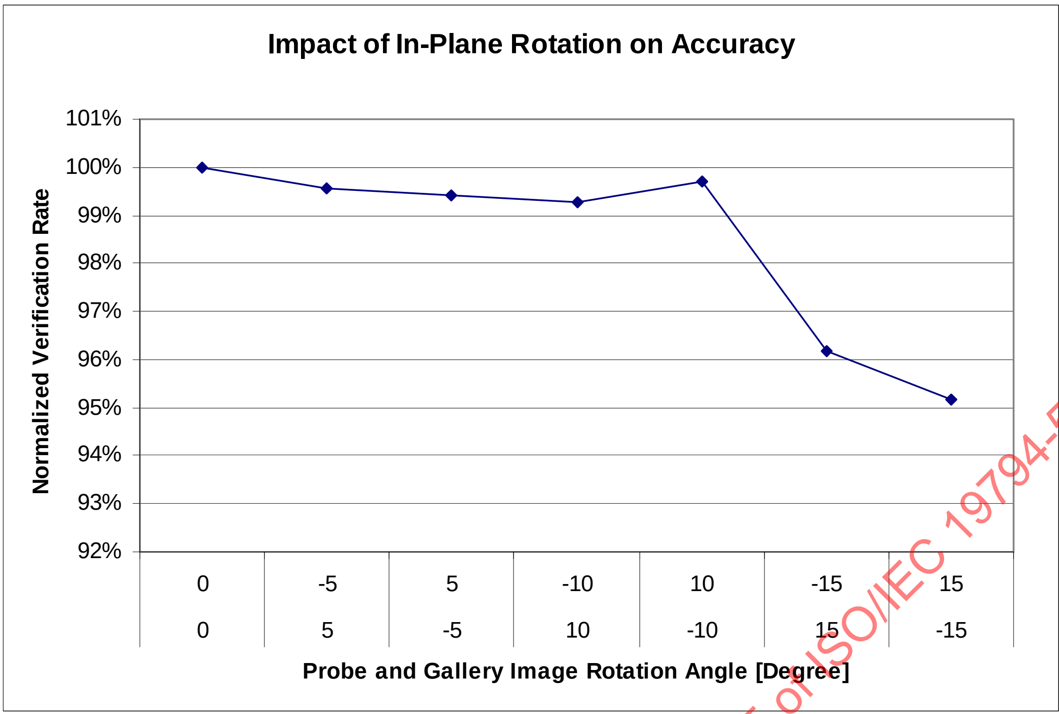


Figure A.16: Impact of in-plane (roll) rotation on automated facial recognition performance: relative verification rate at FAR = 0.1% vs. roll angle

Figure A.17: Sample images to visualize the in-plane rotation as used in the analysis. The face has a roll angle of 0°, +5°, -5°, +10° ,-10°, +15°, -15° (from left to right).

A.7 Best Practices for the Full Frontal 3D Image Type

Besides the requirements of Clause 11 it is best practice for a *Full Frontal 3D Image* to fulfil the following requirements.

A.7.1 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.7.2 Compatibility considerations

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.7.3 Pose of the 3D representation

The rotation of the head should be less than +/- 5° from frontal in roll (ref. Clause 5.5.8).

A.7.4 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.7.5 3D Acquisition Time

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

A.7.6 Best Practices for Full Frontal 3D Image Types using the Range Image Representation

A.7.6.1 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.7.6.2 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 *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 5.6.4.

A.7.7 Best Practices for the Full Frontal 3D Image Types using the 3D Point Map Representation

A.7.7.1 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.7.7.2 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 5.6.4.

A.7.8 Best Practices for Full Frontal 3D Image Types using the 3D Vertex Representation

A.7.8.1 Face coverage

As a best practice requirement for *3D Image Types* using the *Vertex* representation a minimum of 1500 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 5.6.4. 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.8 Best Practices for Token Frontal 3D Images

Besides the requirements of Clause 12 it is best practice for a *Frontal 3D Image* associated with a *Token Frontal Image* to fulfil the following requirements.

A.8.1 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.8.2 Compatibility considerations

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.8.3 Pose of the 3D representation

The rotation of the head should be less than +/- 5° from frontal in roll (ref. Clause 5.5.8).

A.8.4 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.8.5 3D Acquisition Time

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

A.8.6 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.8.7 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.8.8 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.9 Summary of mandatory and best practices for the 3D Image Types

Table A.5 — 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	Supported	Not supported, fixed	Not supported, fixed scaling

coordinate transformation		scaling and offset	and offset
Restrictions on pose (yaw, pitch, roll)	None	None	None

Table A.6 — 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°)

Table A.7 — 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°)

Table A.8 — 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°)

Table A.9 — 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 for Face Image Data

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. 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 document. This annex may also be appropriate source material to application developers or application profile standard developers.

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.

B.2 Photography recommendations

B.2.1 General

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 the requirements of this document and are without shadows or hot spots on the face or excessive glare in eyeglasses.

B.2.2 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 Clause provides expert guidance for the owners and operators of such facilities when they must produce photographs compliant with the requirements of this document.

B.2.2.1 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 10cm, 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 B.1 illustrates these dimensions.

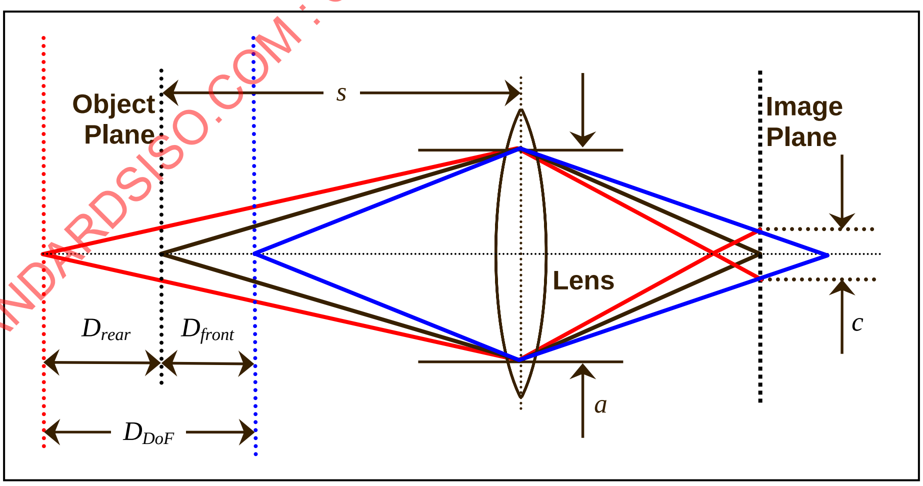


Figure B.1 — Dimensions for depth-of-field calculations

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.

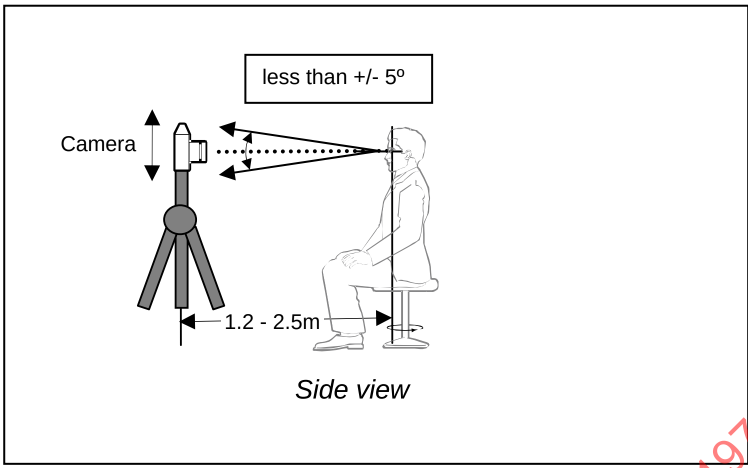


Figure B.2 — Preferred distance and alignment of camera and subject

B.2.2.2 Example of exposure metering at various spots on a subject

Figure B.3 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.

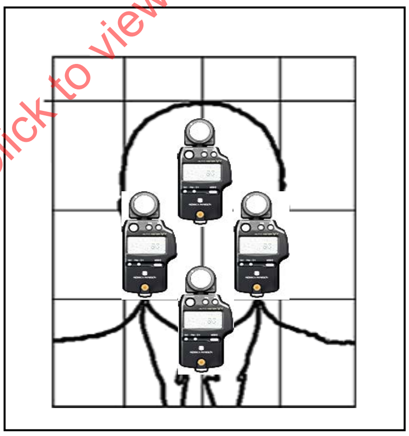


Figure B.3 — Positions of incident light meter for exposure value measurement

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.2.2.3 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.2.2.3.1 Example 1: Proper lighting arrangement with a single light

In this arrangement, illustrated in Figure B.4, 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.

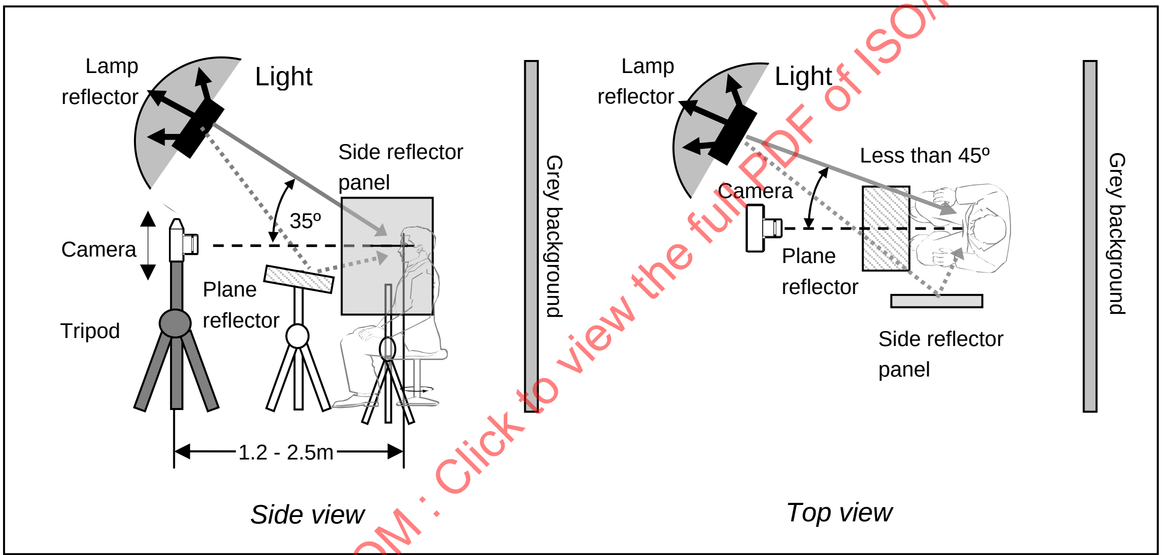


Figure B.4 — Lighting arrangement for a photo studio with a single front light

B.2.2.3.2 Example 2: Proper lighting with dual lights

In the second example illustrated in Figure B.5, 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.

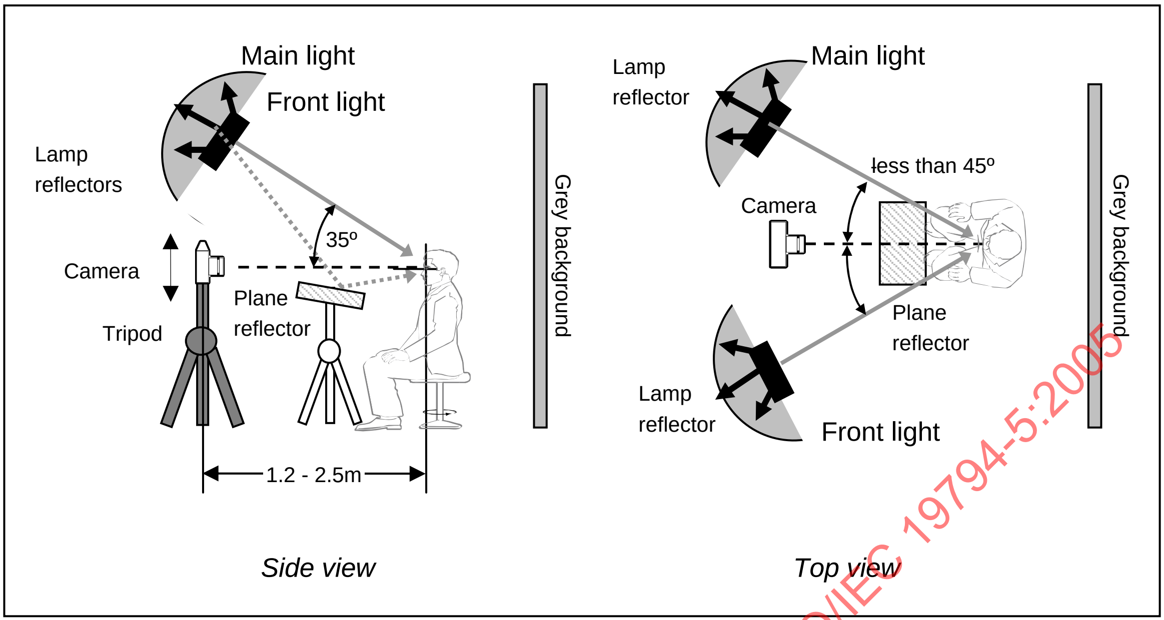


Figure B.5 — Lighting arrangement for a photo studio with dual front lights

B.2.2.3.3 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 B.6, the background light should be aimed at the background and be placed directly behind and below the subject.

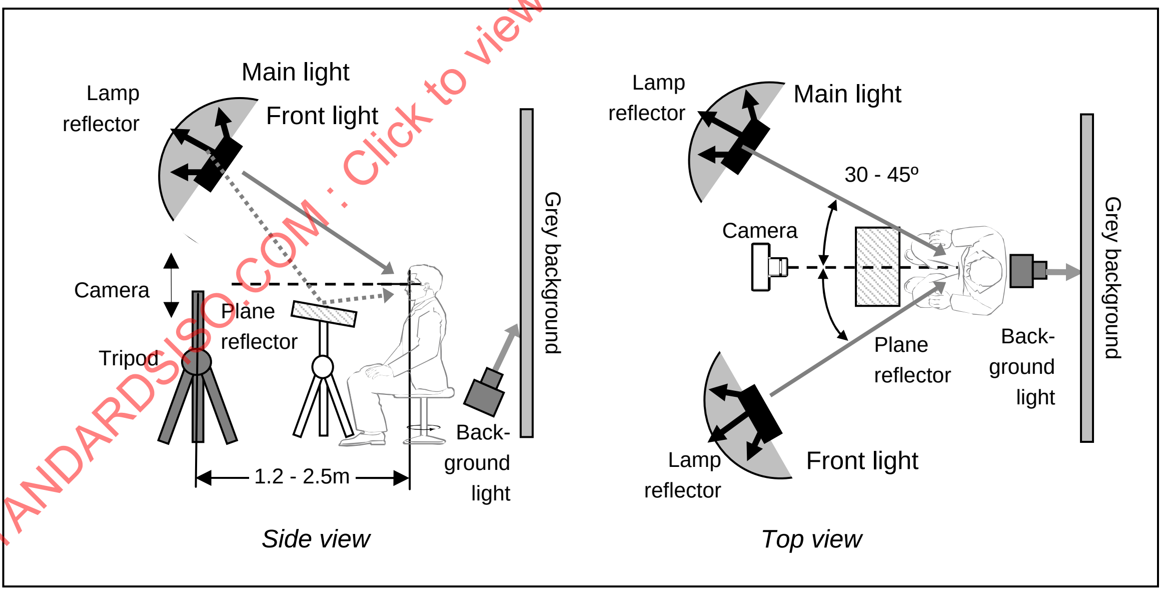


Figure B.6 — Lighting arrangement for a photo studio with dual front lights and a background light

B.2.3 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.2.3.1 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.2.3.2 Example configuration for a photo booth

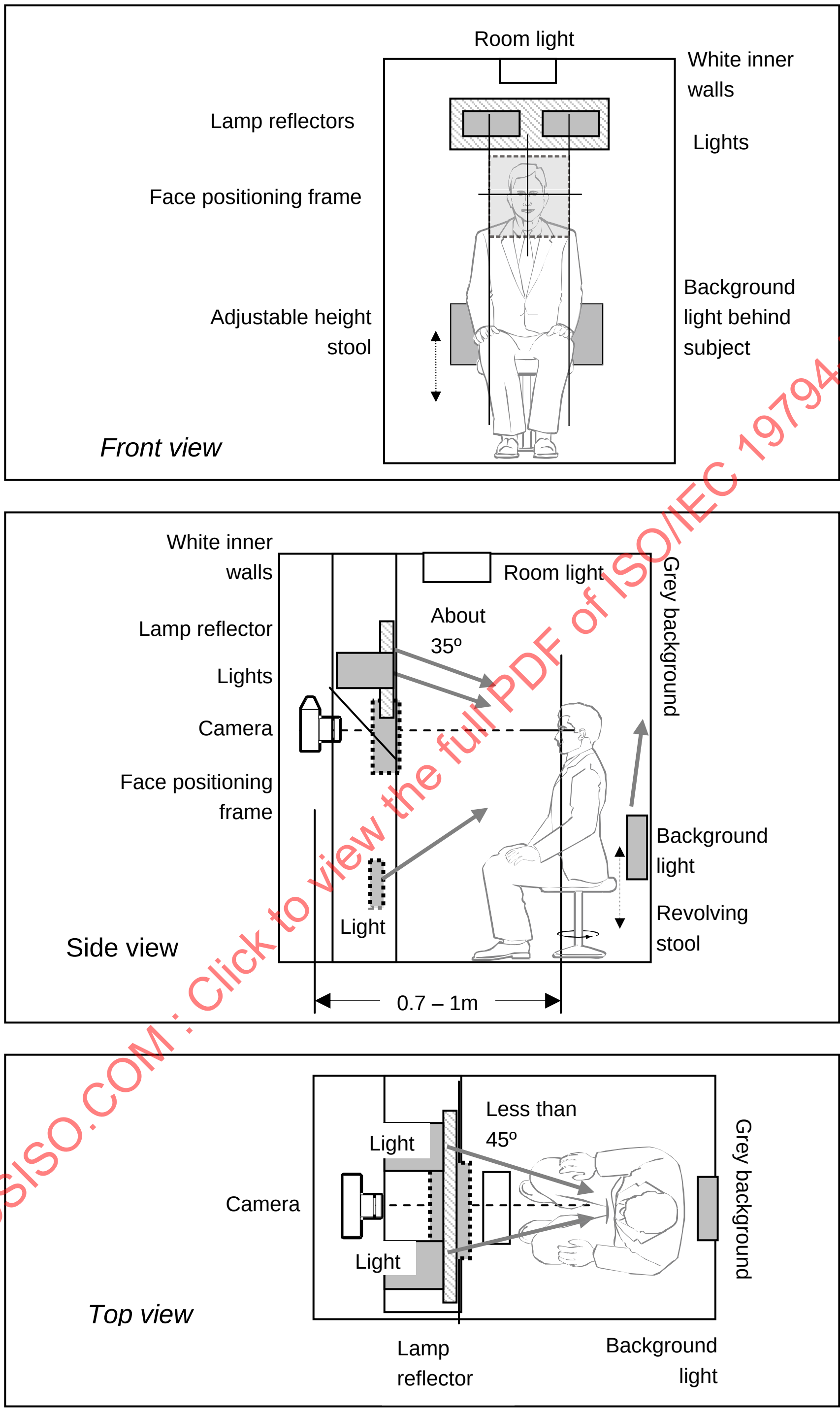


Figure B.7 — Recommended placements of subject, camera, and lights in a photo booth

B.2.3.3 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 centered 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. The camera-to-subject distance should generally be within 0.7-1.0 m.

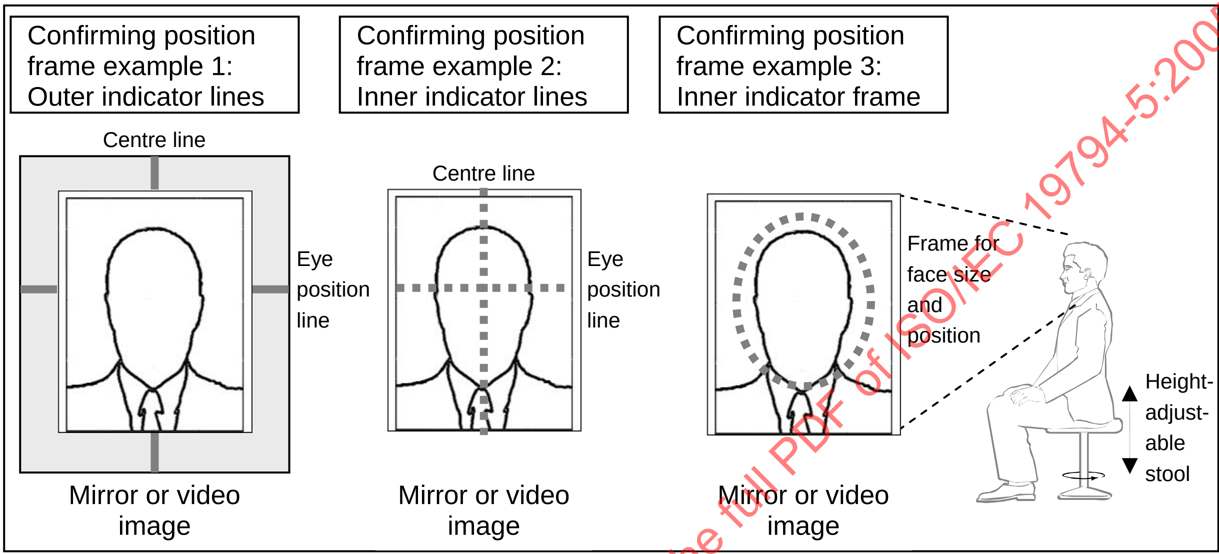


Figure B.8 — Use of a display frame for head positioning

B.2.3.4 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 B.9.

Alternatively, face detection software that automatically sizes and centers 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.