INCITS/ISO/IEC 19794-5:2005[2007]
(ISO/IEC 19794-5:2005, IDT)

American National Standard

INCITS/ISO/IEC 19794-5:2005[2007]
(ISO/IEC 19794-5:2005, IDT)

Information technology — Biometric data interchange formats — Part 5: Face image data

Developed by

incits SM

Where IT all begins

Approved American National Standard
ANSI

Adopted by INCITS (InterNational Committee for Information Technology Standards) as an American National Standard.

Date of ANSI Approval: 8/17/2007

Published by American National Standards Institute,
25 West 43rd Street, New York, New York 10036

Contents

Page

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. 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 national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national 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.

ISO/IEC 19794-5 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 37, *Biometrics*.

ISO/IEC 19794 consists of the following parts, under the general title *Information technology — Biometric data interchange formats*:

— *Part 1: Framework*

— *Part 2: Finger minutiae data*

— *Part 3: Finger pattern spectral data*

— *Part 4: Finger image data*

— *Part 5: Face image data*

— *Part 6: Iris image data*

The following parts are under preparation:

— *Part 7: Signature/sign behavioral data*

— *Part 8: Finger pattern skeletal data*

Introduction

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

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

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

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

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

4) 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 part of ISO/IEC 19794 specifies not only a data format, but also scene constraints (lighting, pose, expression etc), photographic properties (positioning, camera focus etc), digital image attributes (image resolution, image size etc).

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

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

Information technology — Biometric data interchange formats —

Part 5: Face image data

1 Scope

This part of ISO/IEC 19794

— specifies a record format for storing, recording, and transmitting the information from one or more facial images within a CBEFF data structure,

— specifies scene constraints of the facial images,

— specifies photographic properties of the facial images,

— specifies 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 “mugshot” images.

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

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 part of ISO/IEC 19794.

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

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

2 Compliance

Conformity with this part of ISO/IEC 19794 requires compliance with the record format specification defined in clauses 5 and the Basic Face Image Type defined in clause 6.

In addition, this part of ISO/IEC 19794 defines additional Face Image Types. Compliance with the Full Frontal Face Image Type requires compliance with clauses 5, 6, 7, 8. Conformity with the Token Frontal Image Type requires additional compliance with clause 5, 6, 7, 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 10918 (all parts), *Information technology — Digital compression and coding of continuous-tone still images: Requirements and guidelines*

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

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

ISO/IEC 19785 (all parts), *Information technology — Common biometric exchange formats framework*

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

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

4 Terms and definitions

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

4.1
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
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 though a list of face images to match against an input image(s); one-to-many (1:N) searching

4.16
image
two-dimensional representation that encodes the luminance and texture of an object 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) matching

5 The face record format

5.1 Overview

The face 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 one or more images of a human face. This record is embedded in the biometric data block in a CBEFF compliant structure. The record structure is depicted in Figure 2.

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 and the image data will be encoded using either JPEG or JPEG2000.

When referring to elements of the record format, "field" denotes the minimum element such as Face Image Type and Image Data Type, "block" denotes the group of fields such as Facial Information block or Image Information block, and "record" denotes the image data which consists of the Facial Record Header and one or more Facial Record Data.

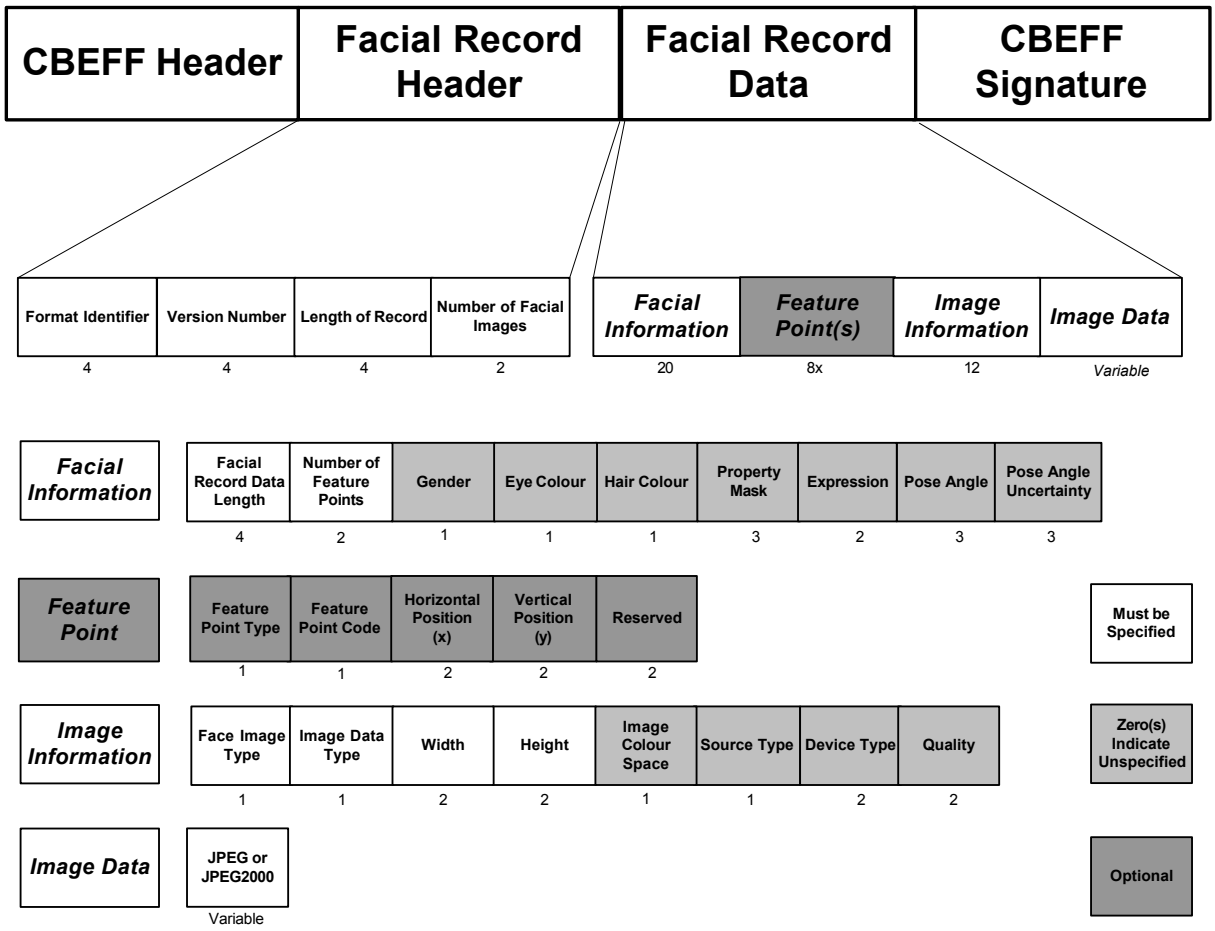


Figure 2 – The Face Image Record format. The length value of each field in bytes is shown below the field. The white boxes indicate fields or blocks that shall be specified, light grey boxes indicate that zero values are used to indicate an unspecified value, and dark grey boxes indicate optional fields

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

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

The organization of the record format is as follows:

- A fixed-length (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 discernable characteristics of the subject such as gender.
    - Multiple (including none) fixed length (8 byte) **Feature Point** blocks describing Feature Points in a facial image.

- 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.

Multiple images of the same person can be described in a single record. This is accomplished by including multiple Facial Record Data blocks after the Facial Record Header block and before the CBEFF Signature block. The structure of this embedding is illustrated in Figure 3.

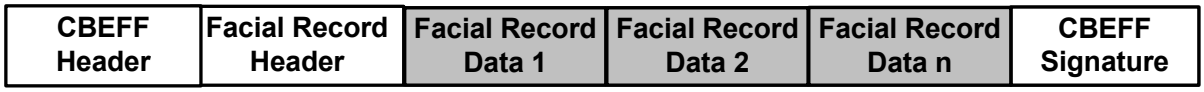


Figure 3 – Embedding multiple images in the same record

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 10^{th} power plus one) would be stored as first byte= 00000100 and second byte=00000001.

5.2.2 Numeric values

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

5.2.3 Conversion to integer

The conversion of a numeric value to integer is given by rounding down if the fractional portion is less than 0.5 and rounding up if the fractional value is greater than or equal to 0.5.

5.2.4 Unspecified field value

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 Part 1 (Framework) of this standard.

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 – The 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	46 < Length of Record ≤ 2^{32} - 1	Includes Facial Record Header and Facial Record Data
Number of Facial Images	2 bytes	1 ≤ Number ≤ 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 of ISO/IEC 19794-5:2004 shall be 0x30313000; “010” – Version 1 revision 0.

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 discernable 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 are different colours, then right eye colour is 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. can not 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; the lowest bit set to 1 shall indicate that all listed properties have been considered and that a zero value of any property bit indicates an absence of that property.

Table 6 – Property flags

Description	Mask Position
Properties are specified	0
Glasses	1
Moustache	2
Beard	3
Teeth visible	4
Blink (either or both eyes closed)	5
Mouth open	6
Left Eye Patch	7
Right Eye Patch	8
Dark Glasses (medical)	9
Feature Distorting Medical Condition (which could impact Feature Point detection)	10
Reserved for future definition	11 – 23

Note that a Blink flag set to “1” will indicate non-compliance with the Frontal, Full Frontal, and Token image types.

5.5.7 Expression

The (2 byte) Expression field shall represent the expression of the face according to Table 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 for future definition	0x00 0x01 – 0x7F	0x08 – 0xFF 0x00 – 0xFF
Reserved for vendor definition	0x80 – 0xFF	0x00 – 0xFF

5.5.8 Pose Angle

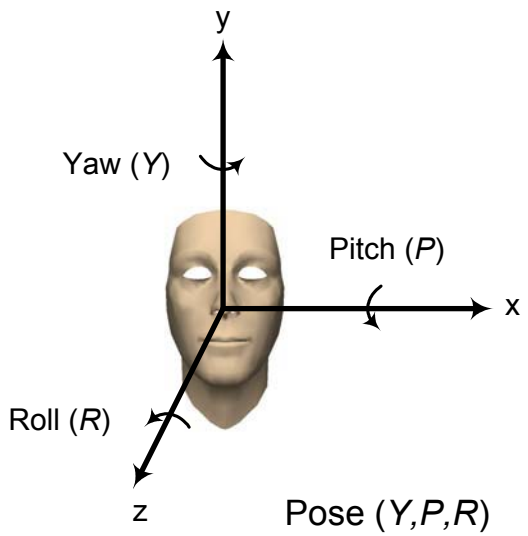


Figure 4 – The definition of pose angles is with respect to the frontal view of the subject

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 4. The examples are shown in Figure 5.

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

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

a **b** **c** **d** **e** **f** **g**

Figure 5 – Examples of pose angles and their encodings. The pose angles (Y, P, R) of figures a – g are given by (0, 0, 0), (+45, 0, 0), (-45, 0, 0), (0, -45, 0), (0, +45, 0), (0, 0, -45), and (0, 0, +45), respectively. The pose angle encodings, (B_Y, B_P, B_R) as defined in sections 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 4. 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 encoding B_Y of the yaw angle Y shall be in degrees as a byte (1 byte) with values from –180 to 180 modulo 2.

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

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

5.5.8.2 Pose Angle – Pitch

The pitch angle P is the rotation in degrees about the x-axis (horizontal axis) shown in Figure 4. Frontal faces have a pitch angle of 0 degrees. Positive angles represent faces looking down (a counter-clockwise rotation around the x-axis).

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

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

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

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

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

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

5.5.9 Pose Angle Uncertainty

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

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

5.6 The Feature Point Block

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

5.6.1 Feature Point Type

The (1 byte) Feature Point Type field represents the type of the Feature Point stored in the Feature Point block. This field shall be set to 0x01 to denote that the position of the Feature Point is represented by the coordinate of the image. All other field values are reserved for future definition of Feature Point types.

5.6.2 Feature Point Code

The (1 byte) Feature Point Code field shall specify the Feature Point that is stored in the Feature Point block. The codes of the Feature Points in clause 5.6.3, taken from the MPEG4 standard and defined as MPEG4 Feature Points, or the additional eye and nostril Feature Points in clause 5.6.4 shall be stored in this block.

Each Feature Point code is represented by a notation A.B using a major (A) and a minor (B) value. The encoding of the Feature Point code is given by the (1 byte) value of A*16 + B.

Table 8 – The Feature Point block

Field	Size	Value	Notes
Feature Point Type	1 byte	0x01	Denotes a 2D Feature Point. All other values are reserved.
Feature Point Code	1 byte	A*16 + B A and B are specified in 5.6.2 and 5.6.3.	The maximum values of A and B are 15.
X coordinate	2 bytes	Horizontal pixel count from upper left pixel.	Count starts at 0.
Y coordinate	2 bytes	Vertical pixel count from upper left pixel.	Count starts at 0.
Reserved	2 bytes	0x00	Reserved for later use with 3D faces.

5.6.3 MPEG4 Feature Points

Figure 6 denotes the Feature Point codes associated with Feature Points as given by Annex C of ISO/IEC 14496-2

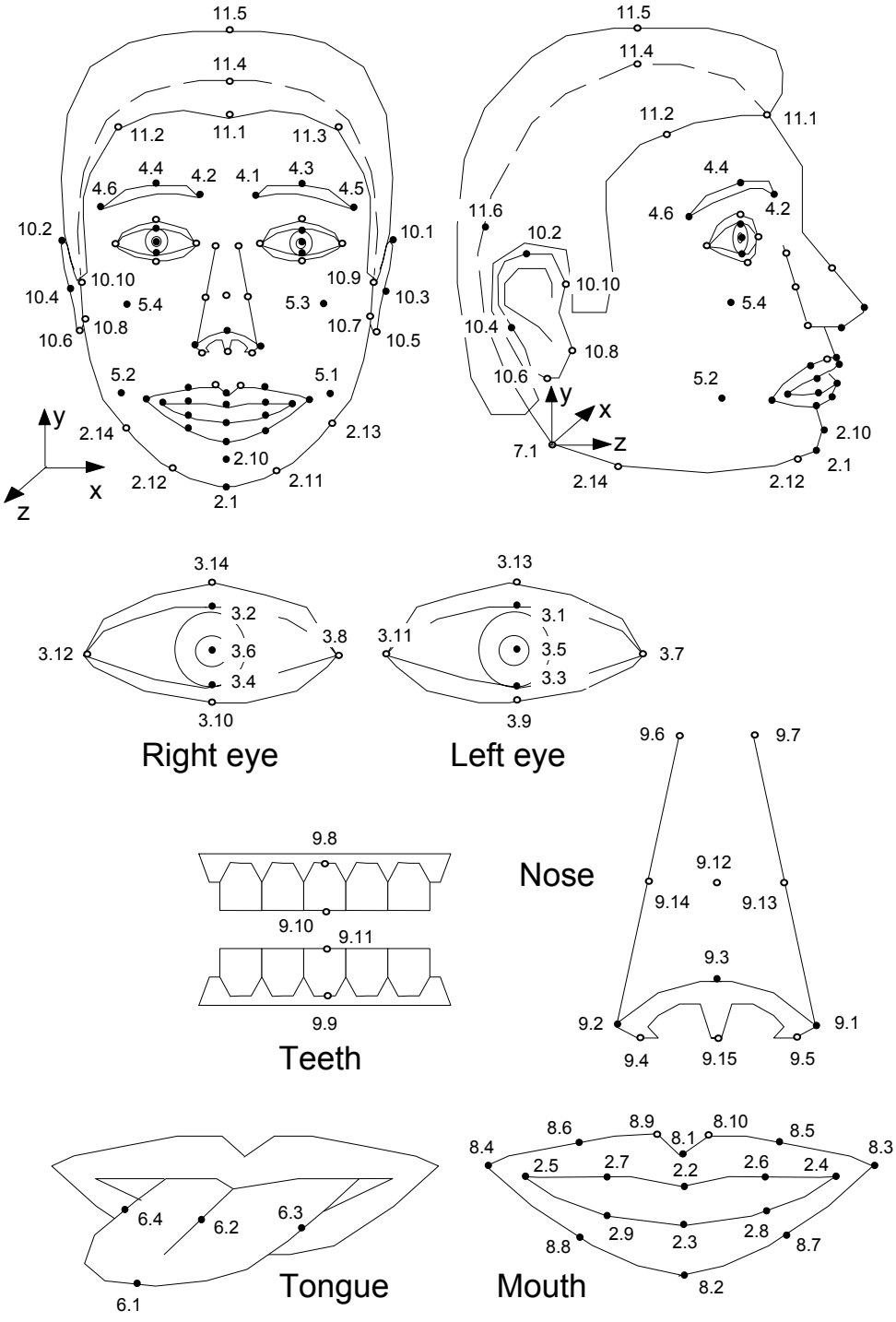


Figure 6 – The Feature Point codes defined in ISO/IEC 14496-2

Each Feature Point code in Figure 6 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 centre Feature Points

The eye centre Feature 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 Feature Point 12.3 is defined to be the midpoint of the nose Feature Points (9.1, 9.15) in the horizontal direction and (9.3,9.15) in the vertical direction. Similarly, the right nostril centre Feature Point 12.4 is defined to be the midpoint of the nose Feature 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 Feature Points are shown in Figure 7 and values given in Table 9.

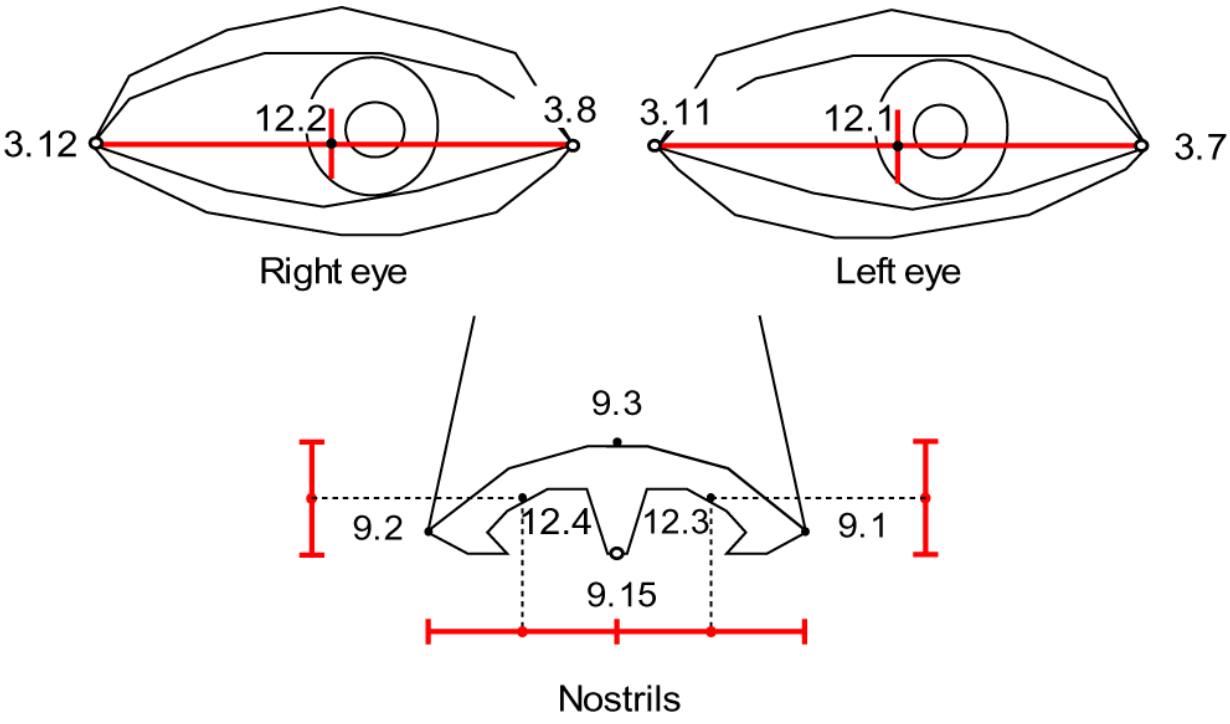


Figure 7 – The eye and nostril centre Feature Points are defined by midpoints of MPEG4 Feature Points

Table 9 – Eye and nostril centre Feature Point codes

Centre Feature Point	Midpoint of Feature Points		Feature 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	

5.7 The Image Information Block

The (16 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 Data Block according to Table 10. Note that all Frontal Image Types are either Full Frontal or Token Frontal. Therefore a separate Frontal Value is not required.

Table 10 – Face Image Type codes

Description	Value
Basic	0x00
Full Frontal	0x01
Token Frontal	0x02
Reserved	0x03 – 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 8.

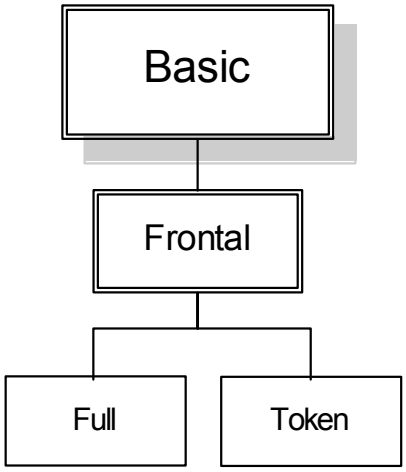


Figure 8 – Face Image Types and their inheritance map. Normative requirements for the Basic, Frontal, Full Frontal and Token Frontal Face Images Types are given in clauses 6, 7, 8, and 9 respectively

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 that a "Unspecified" Value cannot be encoded.

Table 11 – 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 12. The values of 128-255 are vendor specific. Application developers may obtain the values for these codes from the vendor.

Table 12 – 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 13.

Table 13 – 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 zeros 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 14 – Image Data structure

Field	Size	Value	Notes
Image Data	Variable	See Table 11 in clause 5.7.2	Either JPEG or JPEG2000

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 8.

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 (Basic)

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

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

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 “The Basic Face Image Type”. Note that the Frontal Face Image Type is not a valid Face Image Type but helps to describe common specifications of the Full Frontal Image Type and the Token Image Type. Therefore, all Frontal Type Images have to be either Full Frontal or Token Frontal. (Refer 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 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. The full-face frontal pose shall be used. Rotation of the head shall be less than +/- 5 degrees from frontal in every direction – roll, pitch and yaw (ref. 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.

7.2.3 Expression

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

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

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

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

d) Raised eyebrows

e) Eyes looking away from the camera

f) Squinting

g) Frowning

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

7.2.4 Assistance in positioning the face

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

7.2.5 Shoulders

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

7.2.6 Backgrounds

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

7.2.7 Subject and scene lighting

Lighting shall be equally distributed on the face. There shall be no significant direction of the light from the point of view of the photographer, as further described in Clauses 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 section 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 so 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 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 under exposure) on the face.

7.3.3 Focus and depth of field

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

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

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

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

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 “The Basic Face Image Type” and clause 7 “The Frontal Face Image Type”.

8.2 Scene requirements for the Full Frontal Face Image Type

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

8.3 Photographic requirements for the Full Frontal Face Image Type

8.3.1 Introduction

This clause describes the *minimum* relative dimensions of the full image with respect to the face. These minimum requirements will assure that the entire head is in the image, as well as the outline of the shoulders are visible. The requirements of this section can be met by images taken in both portrait and landscape mode, and Figure 9 shows a portrait image and head outline to display dimensions A, B, BB, CC, and DD which are referenced in clauses below.

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

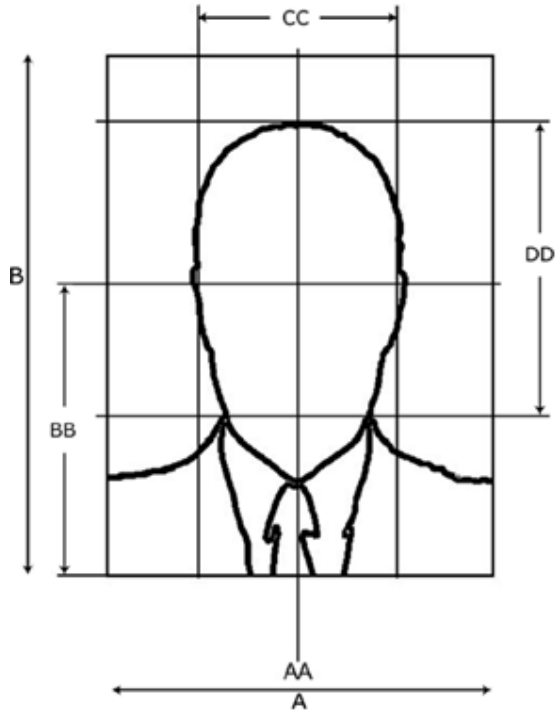


Figure 9 – Geometric characteristics of the Full Frontal Face image

8.3.2 Horizontally centred face

The approximate horizontal midpoints of the mouth and of the bridge of the nose shall lie on an imaginary vertical line AA positioned at the horizontal centre of the image.

8.3.3 Vertical position of the face

The vertical distance BB, denoting the vertical distance from the bottom edge of the image of an imaginary horizontal line passing through the centre of the eyes, shall be between 50% and 70% of the total vertical length B of the image. A single exception is allowed for children under the age or 11 years, in which case the lower limit shall be modified to 40%.

8.3.4 Width of head

The width of a head is defined as the horizontal distance between the midpoints of two imaginary vertical lines; 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 9[1)].

In order to assure that the entire face is visible in the image, the minimum image width shall be specified by the (Image Width: Head Width) ratio (A:CC) of 7:5.

1) This figure is a derivative of AAMVA document DL/ID-2000.

8.3.5 Length of head

The length of a head is defined as the vertical distance between the base of the chin and the crown, shown as length DD in Figure 9.

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 no more than 80% of the vertical length of the image (B).

8.3.6 Summary of photographic requirements

Table15 below summarizes the photographic requirements for full frontal images specified in sections 8.3.3 – 8.3.5.

Table 15: Summary of photographic requirements for Full Frontal Images

Section	Definition	Requirements
8.3.3	Vertical Position of Face	$0.5\ B \leq BB \leq 0.7\ B$
8.3.3	Vertical Position of Face (Children under the age of 11)	$0.4\ B \leq BB \leq 0.7\ B$
8.3.4	Width of Head	$A \geq 1.4\ CC$
8.3.5	Length of Head	$B \geq 1.25\ DD$

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 informative annex section 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 2.

9 The Token Face Image Type

9.1 Inheritance requirements for Token Face Image Type

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

9.2 Digital requirements for the Token Face Image Type

9.2.1 Introduction

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

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

9.2.2 Eye positions

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

1 - to use computer inspection

2 - to use human visual inspection

3 - to use computer and human visual inspection.

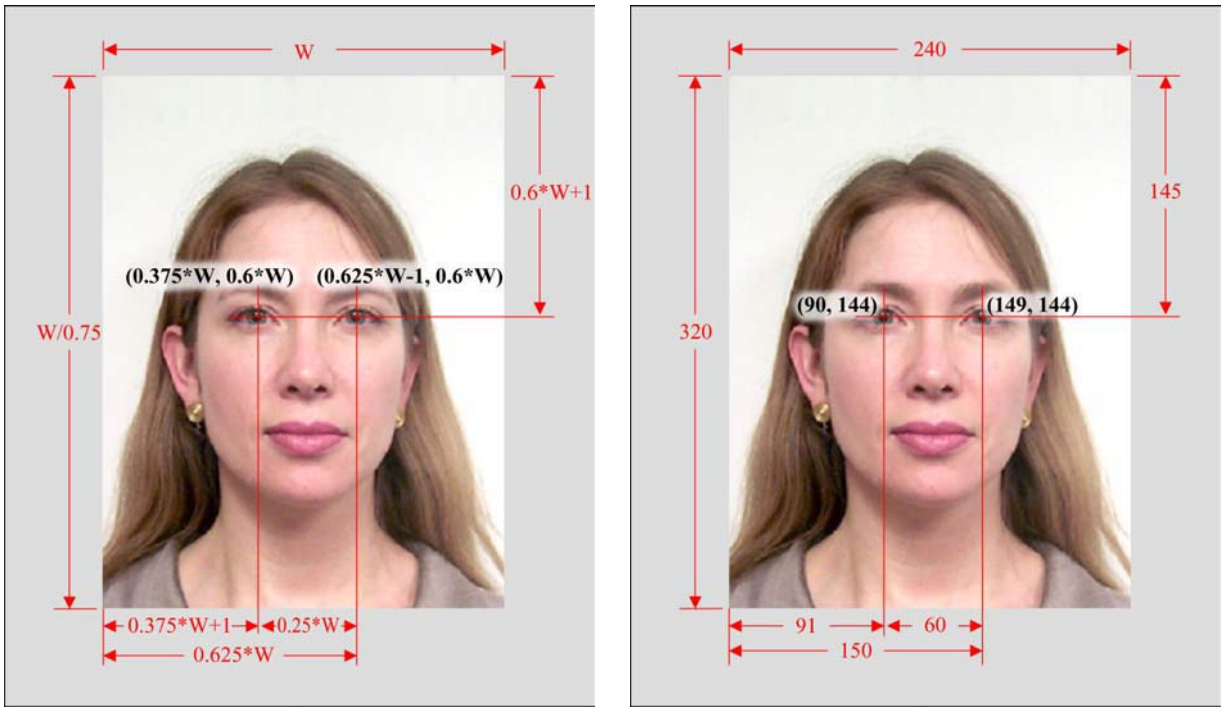
9.2.3 Token image geometric format

A Token image is a colour or greyscale image with image dimensions and eye position coordinates given by Table 16. Note that clause 5.2.3 specifies conversion of values to integer.

Table 16 – The geometric characteristics of the Token Image Type

Feature or Parameter	Value
Image Width	W
Image Height	W/0.75
Y coordinate of Eyes	0.6 * W
X coordinate of First (right) Eye	0.375 * W
X coordinate of Second (left) Eye = 0.625 * W	(0.625 * W) - 1
Width from eye to eye (inclusive)	0.25 *W

An example is shown in Figure 10.



a) b)

Figure 10(a-b) – The Token Face Image Type geometric format (a) and a sample minimum width (W=240) Token Face Image format image (b)

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 10.

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

9.2.5 Padding

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

9.3 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.

Annex A
(informative)

Best practices for Face Images

A.1 Basic Face Images

A.1.1 Purpose

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

A.1.2 Feature Point determination

The Feature Point block defined in section 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 Frontal Images

A.2.1 Purpose

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

A.2.2 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.

A.2.2.1 Examples of unacceptable expressions

a) Closed eyes

b) Hair covering eyes

c) Rim of glasses covering part of the eye

A.2.2.2 Examples of non-recommended expressions

a) A smile where the inside of the mouth and/or teeth is exposed (jaw open)

b) Raised eyebrows

c) Eyes looking away from the camera

d) Squinting

e) Frowning

A.2.3 Assistance in positioning the face

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

A.2.4 Backgrounds

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

A.2.4.1 Background segmentation

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

A.2.4.2 Background shadows

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

A.2.4.3 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.2.4.4 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.2.5 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.2.6 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.2.7 Colour calibration

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

A.2.8 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 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.

1) For a photograph with head width 20mm (roughly 0.78 inches), the recommended scanner resolution is 120 dots per centimetre (roughly 300 dots per inch).

2) For a photograph with head width 13 mm inches (roughly 0.5 inches), the recommended scanner resolution is 189 dots per centimetre (roughly 480 dots per inch).

3) For a photograph with head height (from chin to crown) of 25mm (roughly 1 inch) , this in turn corresponds to a head width on average of roughly 20 mm (roughly 0.8 inches) using a typical head geometric ratio of 4 to 5. This corresponds to a required scanner 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 300dpi.

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 driver's 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 17 summarizes the geometric constraints in clauses A.3.2.1-A.3.2.2.

Table 17: Summary best practices for Full Frontal Images on travel documents

Section	Definition	Recommendation
A.3.2.1	Width to Height Ratio of Image	$1.25 \leq B/A \leq 1.34$
A.3.2.2	Width of Head	$1.4\ CC \leq A \leq 2\ CC$
A.3.2.2	Length of Head	$0.7\ B \leq DD \leq 0.8\ B$
A.3.2.2	Length of Head (Children under the age of 11)	$0.5\ B \leq DD \leq 0.8\ B$

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

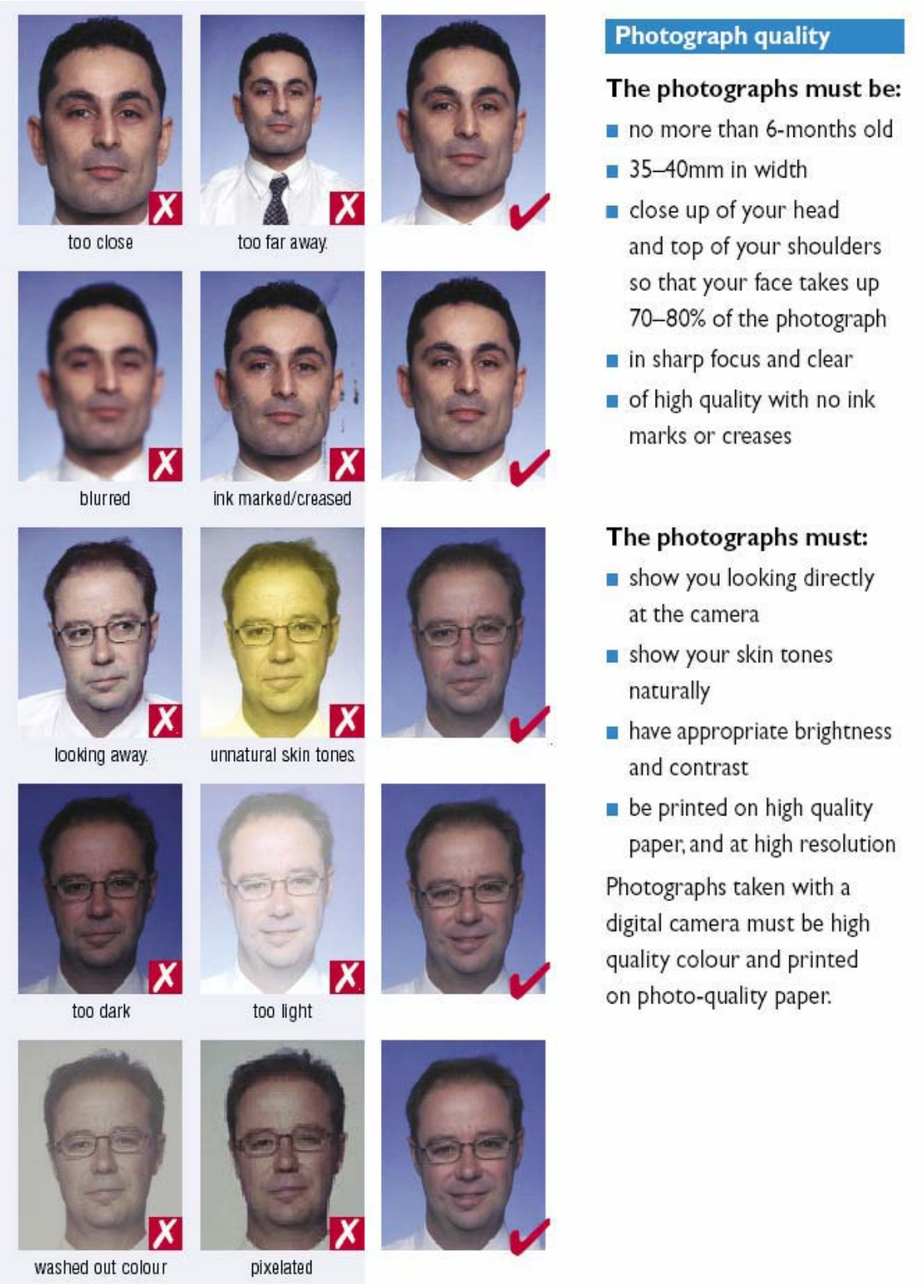


Figure 11 (45 images) – Best practices for the purpose of travel document creation

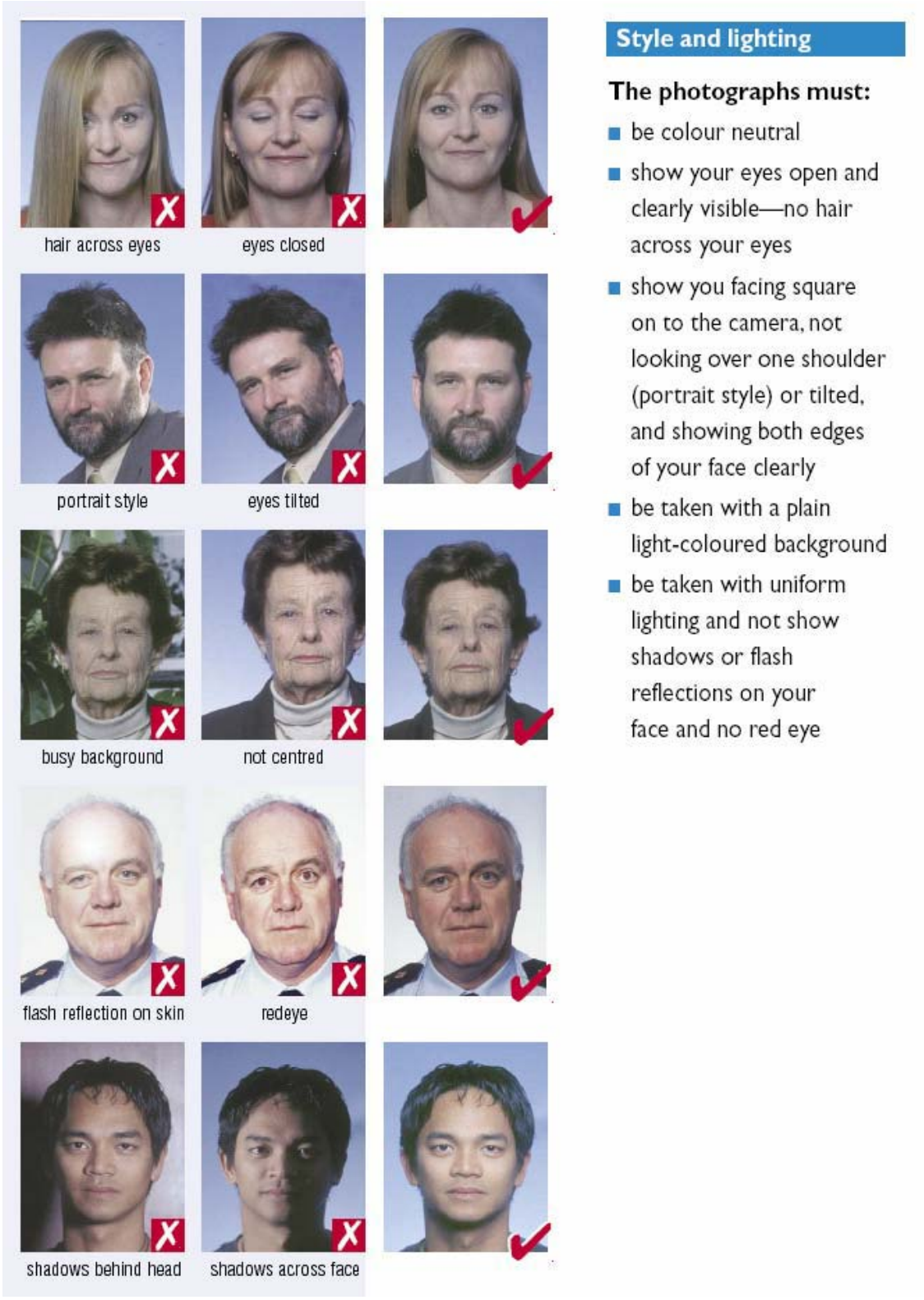


Figure 11 (45 images) – Best practices for the purpose of travel document creation

Glasses and head covers

If you wear glasses:

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

Head coverings:

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

Expression and frame

Your photographs must:

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

Figure 11 (45 images) – Best practices for the purpose of travel document creation

A.3.3 Full Frontal Image Compression

A.3.3.1 Compression – no region of interest

Face recognition performance results for the compression of Full Frontal Images are shown in Figures 12 and 13 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.

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 Section 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.

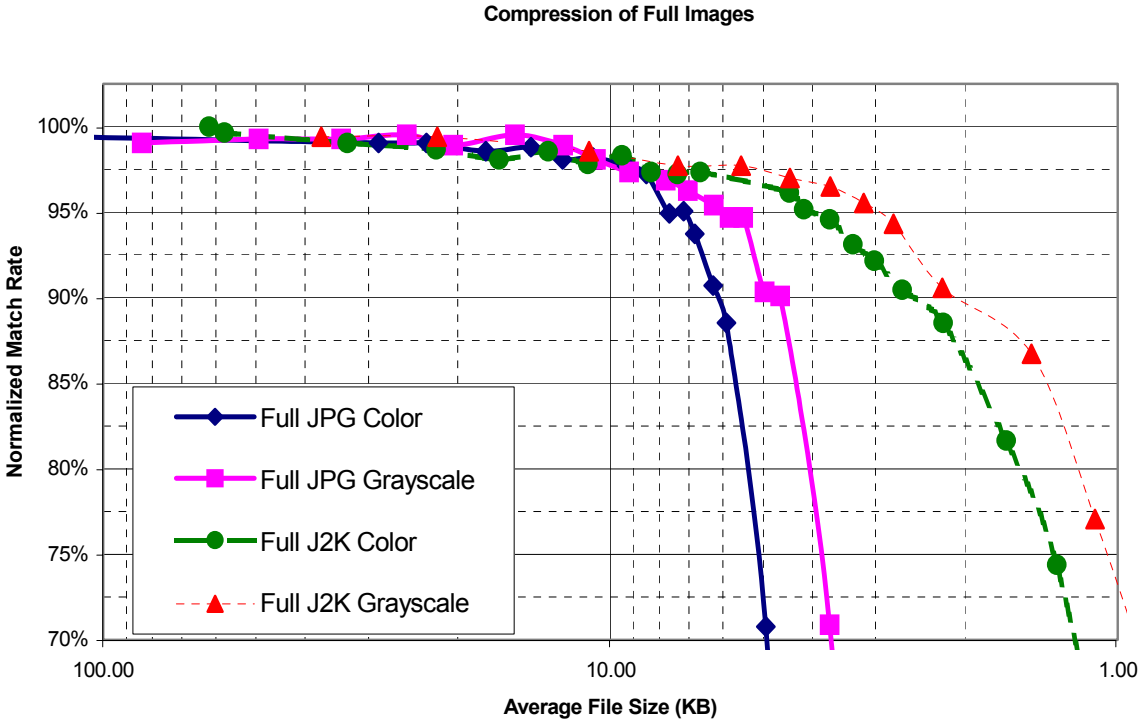


Figure 12 – Identix, compression of Full Frontal Images versus face recognition performance.

2) Provided by Terry Hartmann, Passports Australia, September 2002.

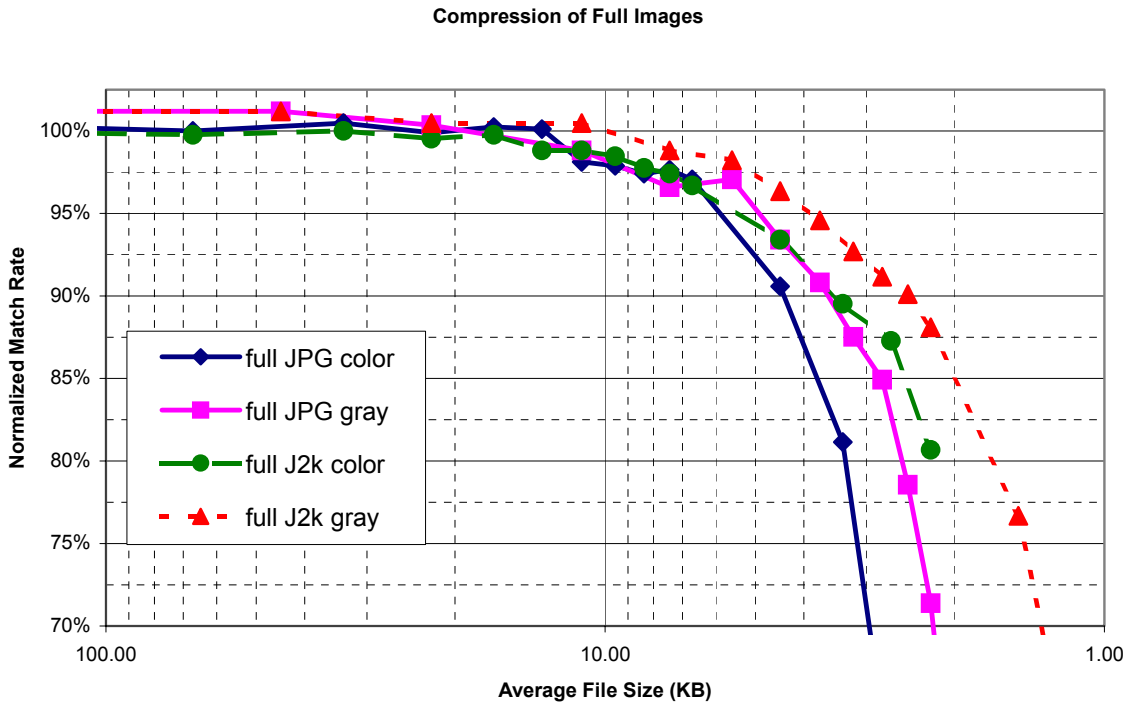


Figure 13 – ZN Vision Technologies, compression of Full Frontal Images versus face recognition performance

In Figures 12 and 13, 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 12 and 13 show that:

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

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

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

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

In conclusion, for use of these two technologies for automatic face 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

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

Example

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

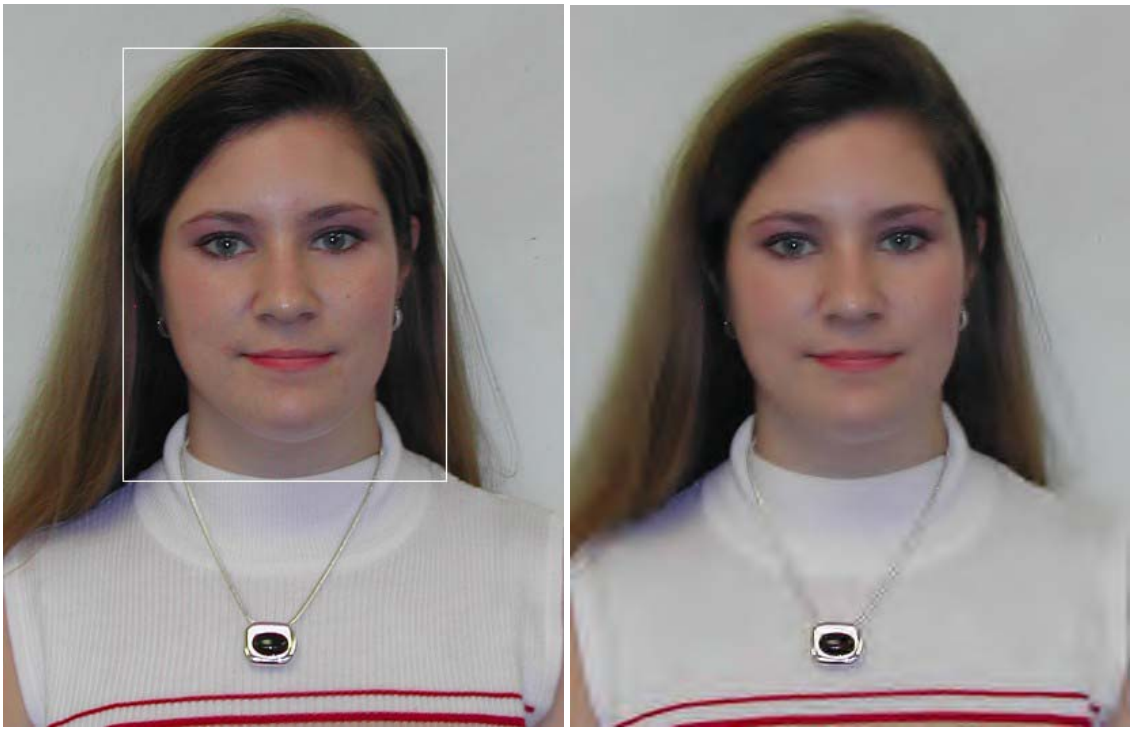
Analysis above indicates that a compression ratio of 60:1 using JPEG2000 preserves matching performance. If a 50:1 ratio is used for the Inner Region, and 200:1 can be used on the Outer Region with an acceptable level of degradation for visual inspection purposes. For a colour, 300 dpi, 35x45mm JPEG2000 image (413x531 pixels, 658 KB uncompressed), with a 240x320 (230.4 KB) Inner Region as defined in Figure 14,

A. 200:1 Outer Region: (658-230.4 KB)/200 = 2.14 KB

B. 50:1 Inner Region: (230.4 KB)/50 = 4.61 KB

Total file size: 2.14 + 4.61 = 6.75 KB. File size reduction: ~40%

In Figure 14 below, the image on the left represents the uncompressed image, and shows the bounds of the Inner Region. The image on the right is compressed using JPEG2000 ROI as described above.



a) b)

Figure 14(a-b) – Example uncompressed (a) and compressed (b) using region of interest shown in (a)

A.4 Token Images

A.4.1 Token image sizes

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

Interpolation required in the affine transformation used to create a Token image can have the effect of introducing artefacts that can harm the face recognition process. For example, if one company chose to use 70 pixels from eye to eye while another chose 60, there might be unnecessary problems. Therefore, in order to improve the interoperability of Token images, it is recommended that the width be specified in units of 240. Examples are given in Table 18.

Table 18 – 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 15 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 Token image (e).

Figure 15 – 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 16 and 17 below, from faces obtained from Passports Australia within the Australian Department of Foreign Affairs and Trade[3)]. Here, 1000 matching pairs (original and renewals) of real passport images were considered.

See section A.2.4 for a detailed description of the compression experiments. The same experiment for Full Frontal images discussed in section A.2.4 was repeated for Token images. The results for Identix and ZN Vision Technologies are shown in Figures 16 and 17.

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

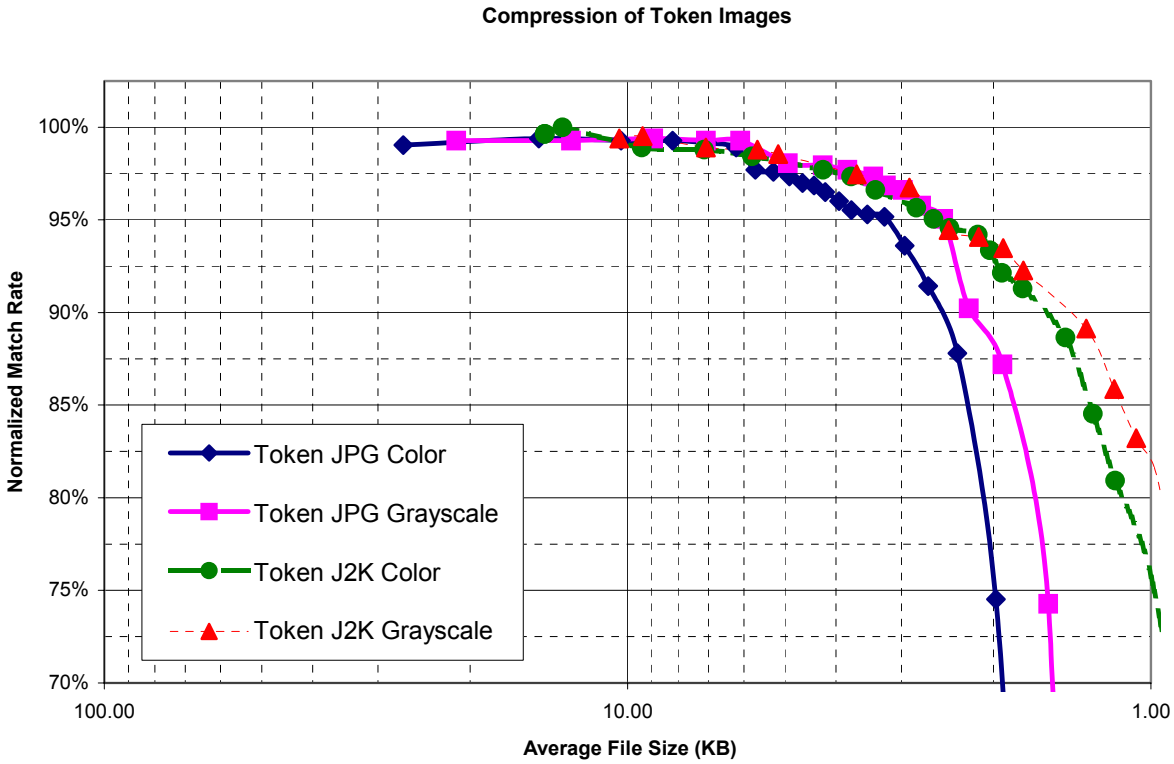


Figure 16 – Identix, compression Token images versus face recognition performance.

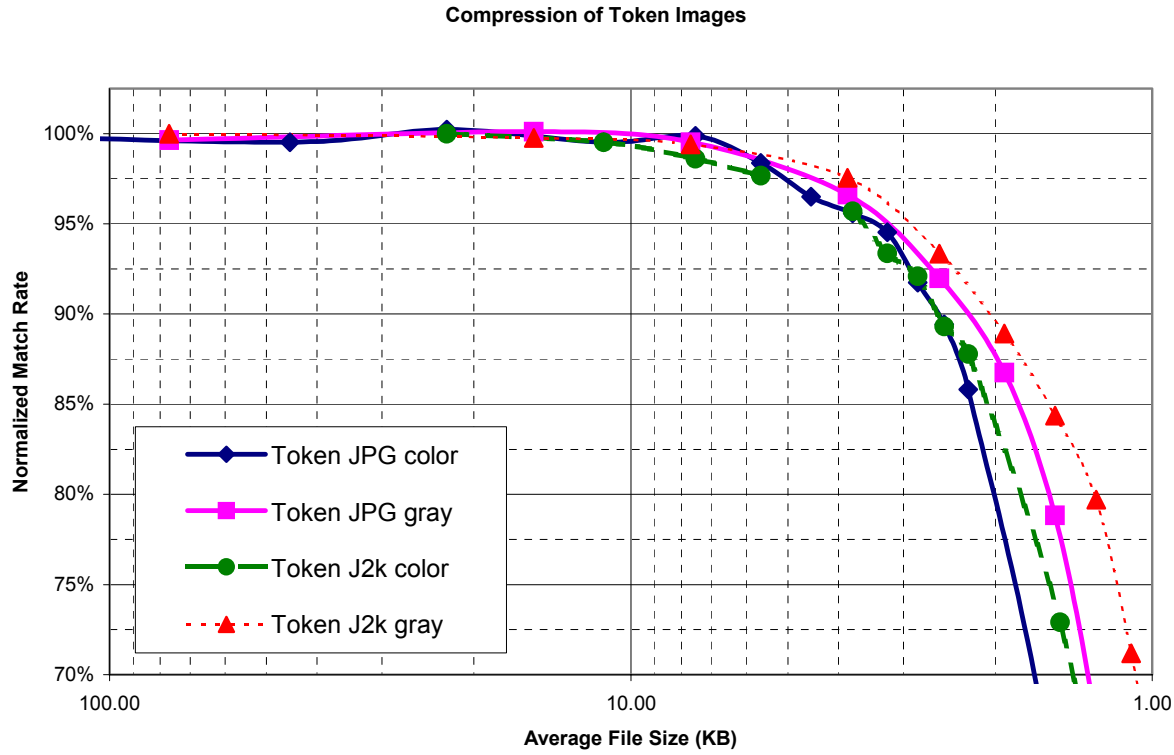


Figure 17 – ZN Vision Technologies, compression of Token Images versus face recognition performance.

The graphs in Figures 16 and 17 show that:

3. The performance degrades quickly below a compressed file size of 8KB.

4. 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 18. This is a region of 192 x 240 pixels in size.

Figure 18 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.

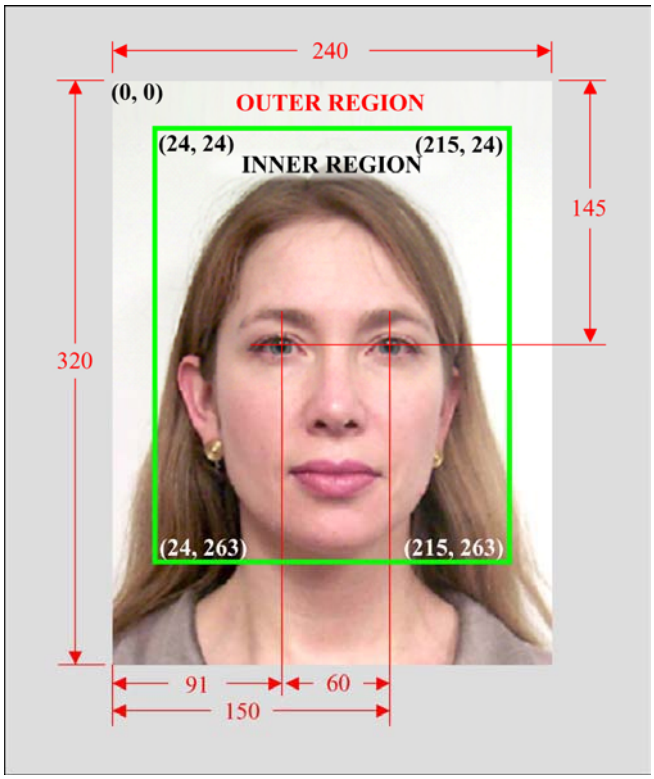


Figure 18 – Suggested region of interest for Token Images

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

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:—[4], *Information technology — Biometric application programme interface — BioAPI specification*

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

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

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

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

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

4) To be published.

ICS 35.040

Price based on 47 pages