INTERNATIONAL STANDARD

ISO/IEC 19794-5

First edition
2005-06-15

AMENDMENT 2
2009-11-01

Information technology — Biometric data interchange formats —

Part 5: Face image data

AMENDMENT 2: Three-dimensional face image data interchange format

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

Partie 5: Données d'image de la face

AMENDEMENT 2: Format d'échange de données d'image de la face tridimensionnelles

Reference number
ISO/IEC 19794-5:2005/Amd.2:2009(E)

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

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.

Amendment 2 to ISO/IEC 19794-5:2005 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 37, *Biometrics*.

This amendment is intended to establish a data interchange format for storing three-dimensional (3D) human face images. To achieve this, several new image types are introduced that are a combination of 2D facial images and associated 3D shape information.

This amendment describes the necessary changes to the data interchange format regarding the capability to hold 3D information and the additional requirements for 3D data.

Information technology — Biometric data interchange formats —

Part 5: Face image data

AMENDMENT 2: Three-dimensional face image data interchange format

Page 3

Add the following reference to Clause 3:

ISO/IEC 15948:2004, *Information technology — Computer graphics and image processing — Portable Network Graphics (PNG): Functional specification*

Page 5

Replace 4.16 with the following:

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

Page 5

Add the following definitions to Clause 4:

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 values of the point in 3D

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 bibliography item [13] for details.

Page 5

Replace clause 5.1 with the following:

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 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.

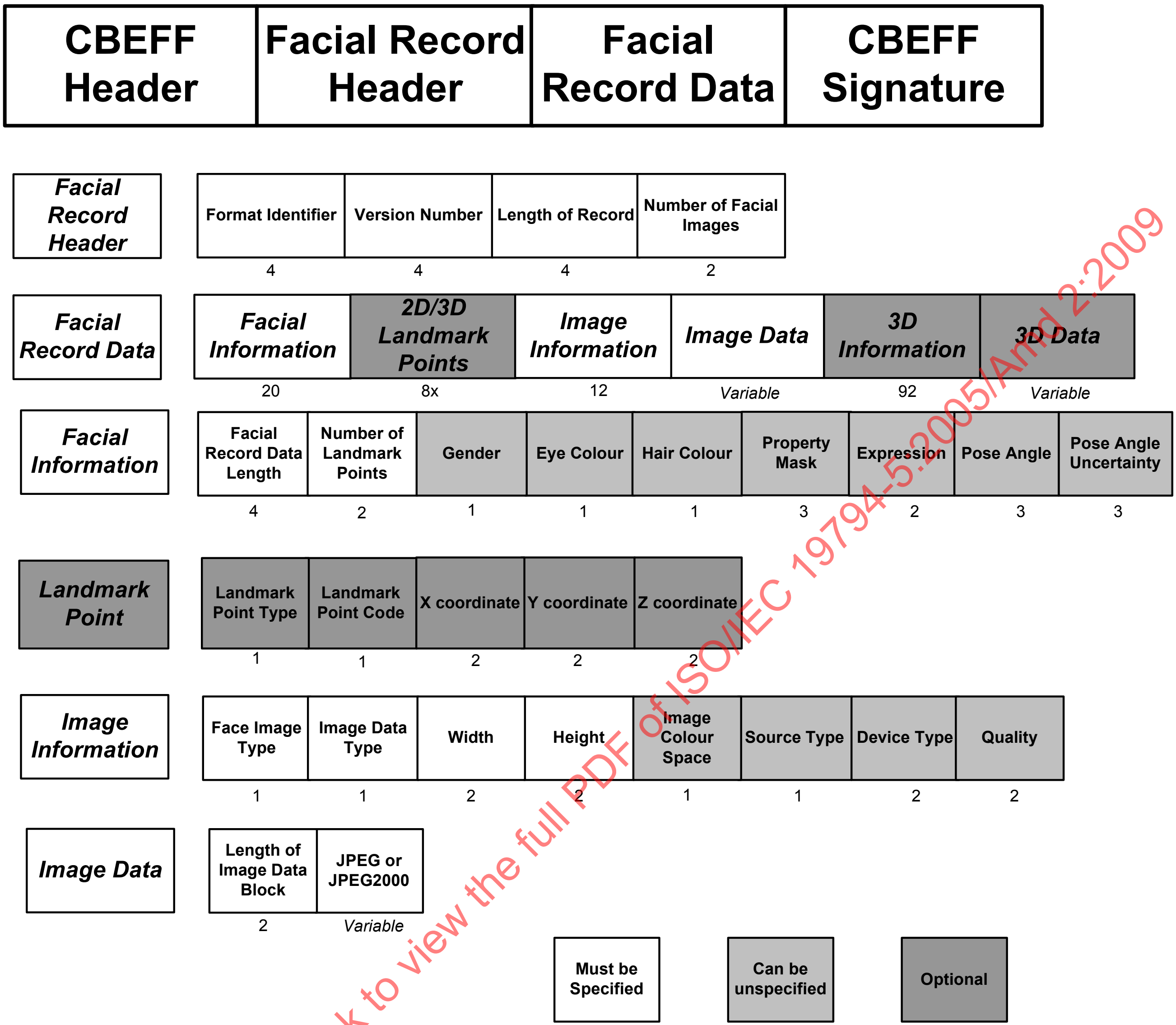

Figure 2 — The Face Image Record format. The length value of each field in bytes is shown below the field. The white boxes indicate fields or blocks that shall be specified, light grey boxes that the fields are mandatory, but an unspecified value is acceptable, and dark grey boxes indicate optional fields. Note, that the 3D Information block and 3D Data block are mandatory for the 3D Types.

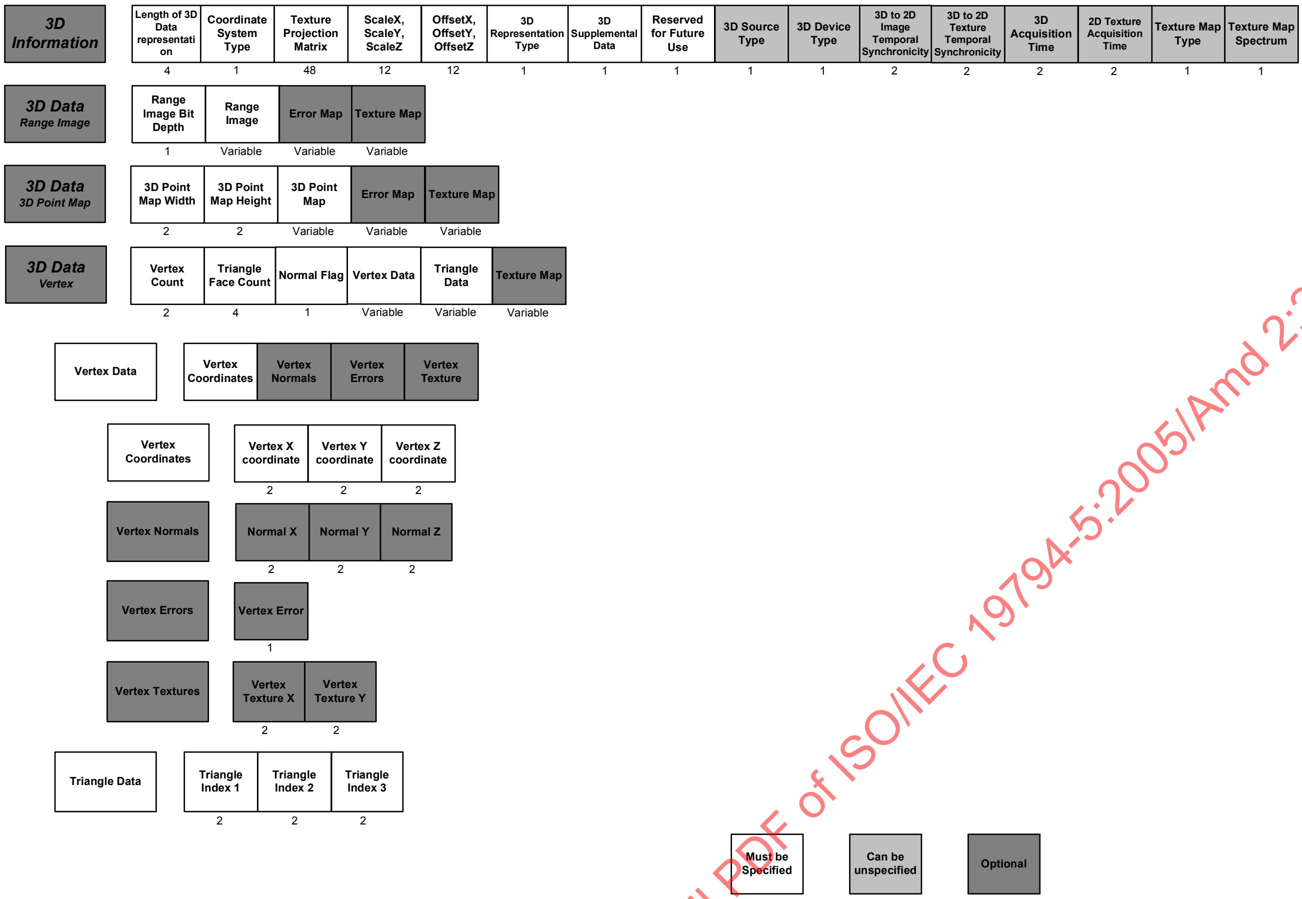

Figure 3 — The 3D Information block and the three possible 3D Data blocks specified in this standard. The length value of each field in bytes is shown below the field. The white boxes indicate fields or blocks that shall be specified, light grey boxes that the fields are mandatory, but an unspecified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

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 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) **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 Amd.2-1.

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 Amd.2-1 – Embedding multiple images / 3D representations in the same record.

Page 8, 5.4

Replace Table 2 with the following:

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	0x30323000 (‘0’ ’ 2’ ’0’ 0x0)	“020” in ASCII
Length of Record	4 bytes	48 < Length of Record ≤ 2^{32} - 1	Includes Facial Record Header and Facial Record Data
Number of Facial Images / 3D representations	2 bytes	1 ≤ Number ≤ 65535	

Page 8

Replace clause 5.4.2 with the following:

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 shall be 0x30323000; “020” – Version 2 revision 0.

Page 15

Replace 5.6 with the following:

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 14496-2:2004 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.

The horizontal and vertical position 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 Amd.2-1	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 Amd.2-1	Denotes the Z-coordinate 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 landmark point is an MPEG4 Feature Point as given by Annex C of ISO/IEC 14496-2:2004 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 it's 3D coordinates. All other field values are reserved for future definition of Landmark Point types.

Table Amd.2-1 — The Landmark Point Type

Description	Value	Comment
MPEG4 Feature	0x01	The Horizontal and Vertical position of the landmark point are measured in pixels with values from 0 to Width-1 and Height-1, respectively. The Z coordinate field is ignored.
Anthropometric 2D landmark	0x02	The landmark point is considered as a 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 14496-2:2004 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 6 denotes the Landmark Point codes associated with Feature Points as given by Annex C of ISO/IEC 14496-2:2004. 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.

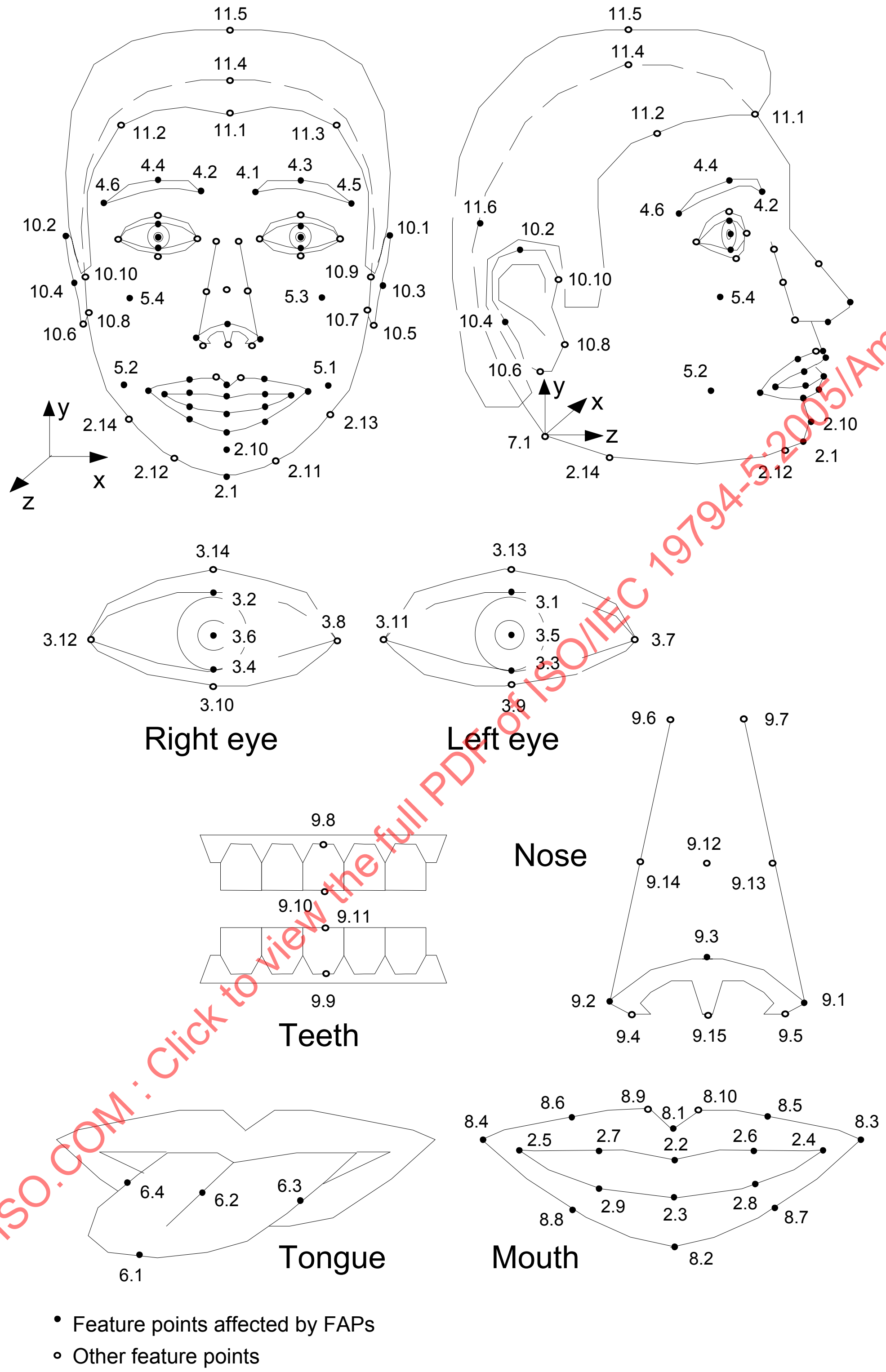

Figure 6 – The MPEG4 Feature Point codes defined in ISO/IEC 14496-2.

Each Landmark Point code in Figure 7 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 7 and values given in Table 9.

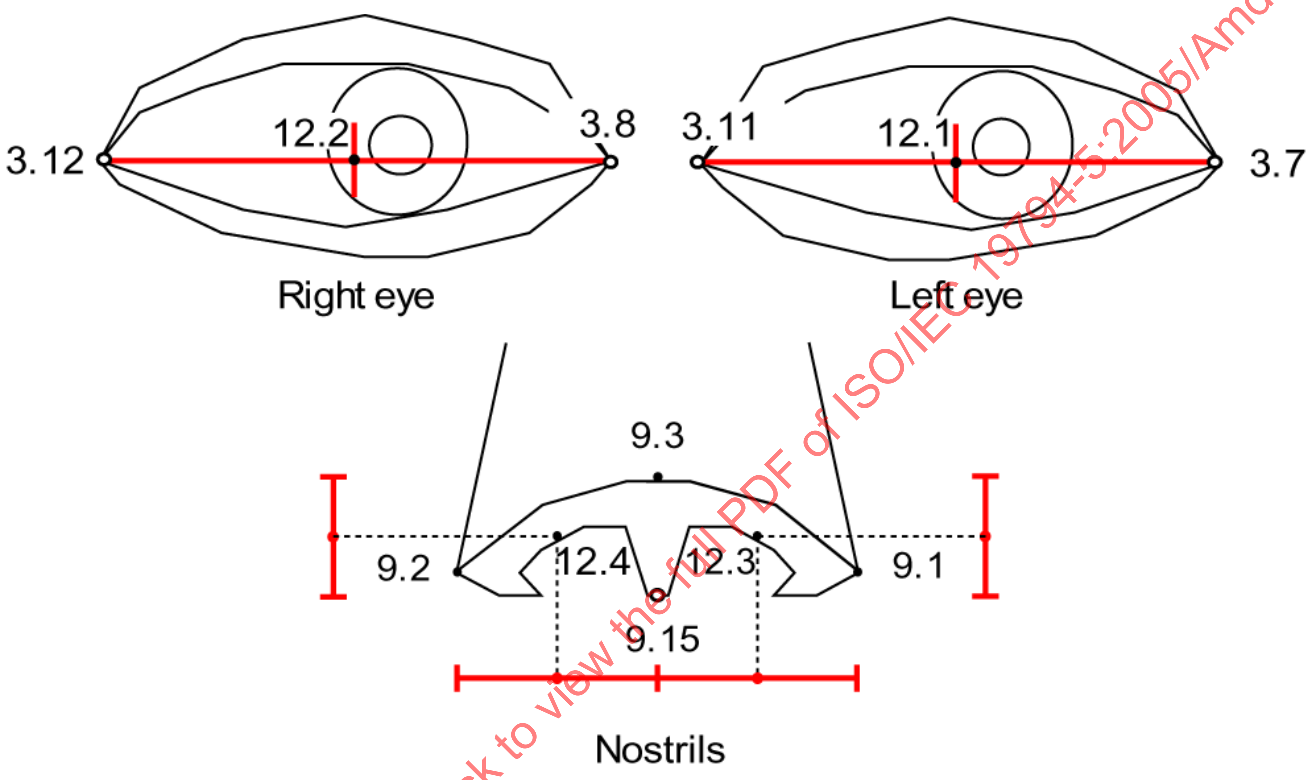

Figure 7 – The eye and nostril centre Landmark Points are defined by midpoints of MPEG4 Feature Points.

Table 9 – 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 Amd.2-2 and Table Amd.2-2 show the definition of the Anthropometric Landmarks. The set of points represents the craniofacial landmark points of the head and face. The latter are used in forensics for “Face to face” and “Skull to face” identification. Some of these points have MPEG 4 counterparts, others not.

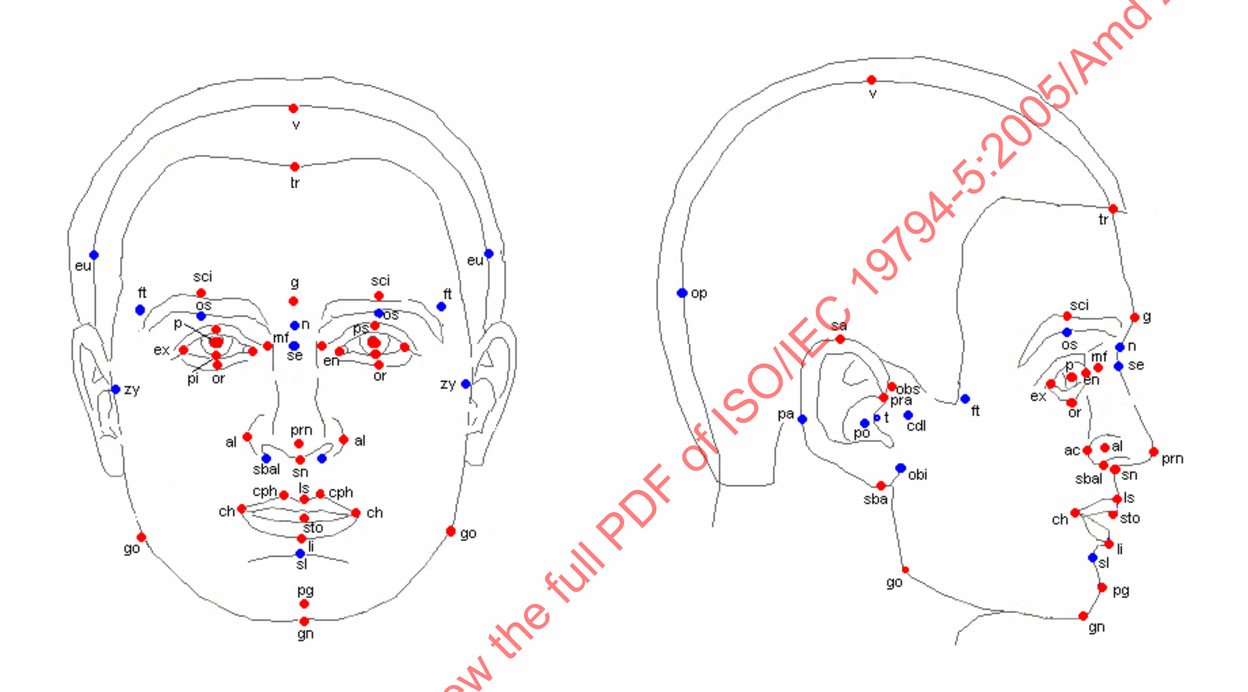

Figure Amd.2-2: Anthropometric Landmarks with (red) and without (blue) MPEG4 counterparts.

Definitions for these points are presented in the Table Amd.2-2.

Table Amd.2-2 – 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 (or gnathion)	The lowest median landmark on the lower border of the mandible
cdl	2.9 2.10		condylion laterale	The most lateral point on the surface of the condyle of the mandible
en	3.1 3.2	3.11 3.8	endocanthion	The point at the inner commissure of the eye fissure
ex	3.3 3.4	3.7 3.12	exocanthion (or ectocanthion)	The point at the outer commissure of the eye fissure
p	3.5 3.6	3.5 3.6	center point of pupil	Is determined when the head is in the rest position and the eye is looking straight forward
or	3.7 3.8	3.9 3.10	orbitale	The lowest point on the lower margin of each orbit
ps	3.9 3.10	3.1 3.2	palpebrale superius	The highest point in the midportion of the free margin of each upper eyelid

pi	3.11 3.12	3.3 3.4	palpebrale inferius	The lowest point in the midportion of the free margin of each lower eyelid
os	4.1 4.2		orbitale superius	The highest point on the lower border of the eyebrow
sci	4.3 4.4	4.3 4.4	superciliare	The highest point on the upper border in the midportion of each eyebrow
n	5.1		nasion	The point in the middle of both the nasal root and nasofrontal suture
se	5.2		sellion (or subnasion)	Is the deepest landmark located on the bottom of the nasofrontal angle
al	5.3 5.4	9.1 9.2	alare	The most lateral point on each alar contour
prn	5.6	9.3	pronasale	The most protruded point of the apex nasi
sn	5.7	9.15	subnasale	The midpoint of the angle at the columella base where the lower border of the nasal septum and the surface of the upper lip meet
sbal	5.9 5.10		subalare	The point at the lower limit of each alar base, where the alar base disappears into the skin of the upper lip
ac	5.11 5.12	9.1 9.2	alar curvature (or alar crest) point	The most lateral point in the curved base line of each ala
mf	5.13 5.14	9.6 9.7	maxillofrontale	The base of the nasal root medially from each endocanthion
cph	6.1 6.2	8.9 8.10	christa philtri landmark	The point on each elevated margin of the philtrum just above the vermilion line
ls	6.3	8.1	labiale (or labrale) superius	The midpoint of the upper vermillion line
li	6.4	8.2	labiale (or labrale) inferius	The midpoint of the lower vermillion line
ch	6.5 6.6	8.3 8.4	cheilion	The point located at each labial commissure

sto	6.7		stomion	The imaginary point at the crossing of the vertical facial midline and the horizontal labial fissure between gently closed lips, with teeth shut in the natural position
sa	7.1 7.2	10.1 10.2	superaurale	The highest point of the free margin of the auricle
sba	7.3 7.4	10.5 10.6	subaurale	The lowest point of the free margin of the ear lobe
pra	7.5 7.6	10.9 10.10	preaurale	The most anterior point on the ear, located just in front of the helix attachment to the head
pa	7.7 7.8		postaurale	The most 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 MPEG4 Feature or Anthropometric 2D landmark. In case the Landmark Point Type equals Anthropometric 3D landmark this field along with the horizontal and vertical positions denotes the coordinates of the landmark point in the 3D Cartesian Coordinate System. The metric coordinates of 3D landmarks shall be obtained by multiplying the X, Y, and Z coordinates by a fixed scale of 0.02 mm. Note, that the Landmark Point Type field codes the type of the landmark point and determines the interpretation of the Z coordinate.

Page 19

Replace 5.7.1 with the following:

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 10. Note that all Frontal Image Types are either Full Frontal, Token Frontal, or one of the respective 3D Full Frontal or Token Frontal Image Types. 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 – 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 8. 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.

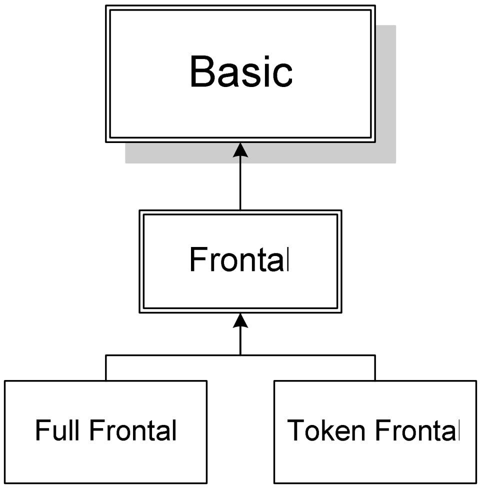

Figure 8 – 2D Face Image Types and their inheritance map. Normative requirements for the Basic, Frontal, Full Frontal and Token Frontal Face Images Types are given in clauses 6, 7, and 8, respectively.

Page 22

Add the following new subclauses after 5.8

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 in metric world coordinates with two disadvantages:

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

Thus the standard features several ways to store 3D data in different representations. All representations support a Cartesian coordinate system. The Range data representation additionally supports a Cylindrical Coordinate system. Note that the coordinate system may be further restricted for different Face Image Types (see clauses 10 to 12, and Annex A.7).

The transformation to metric world coordinates is described by appropriate scaling factors and implicit rules (e.g. as used in the 3D anthropometric landmark type) defined in this standard (see clause 5.9.2.1 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 Amd.2-3 – 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:

$$X = x * ScaleX + OffsetX;$$
$$Y = y * ScaleY + OffsetY;$$
$$Z = z * ScaleZ + OffsetZ.$$

For certain Image Types the origin of the Cartesian Coordinate System shall be the nose, i.e. the Prn landmark as defined in Table Amd.2-2.

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.

Furthermore note, that 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.

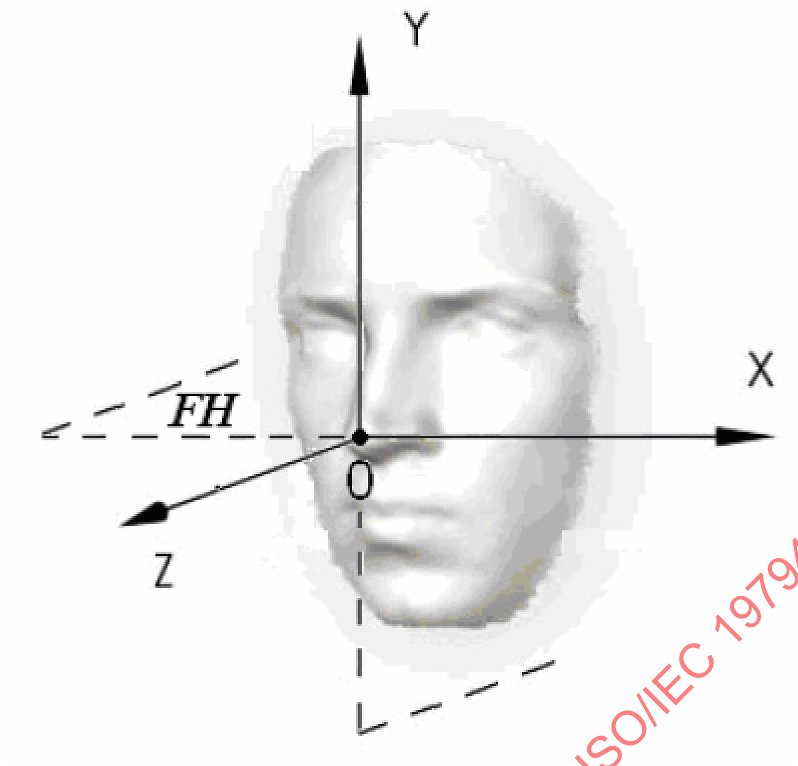

Figure Amd.2-3 – A sample of a Cartesian Coordinate System

5.9.2.2 The Cylindrical Coordinate System

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

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

$$X = r * ScaleZ * \sin(\alpha * ScaleX) + OffsetX;$$
$$Y = h * ScaleY + OffsetY;$$
$$Z = r * ScaleZ * \cos(\alpha * ScaleX) + OffsetZ;$$

ScaleX, ScaleY, ScaleZ, OffsetX, OffsetY and OffsetZ are the necessary constants for the transformation. ScaleX has the physical unit of rad (degree radian). ScaleY, ScaleZ, OffsetX, OffsetY and OffsetZ are given in the physical unit mm (millimetre). Note, that large values of ScaleX, ScaleY or ScaleZ indicate a low 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 landmark as defined in Table Amd.2-2.

Furthermore note, that 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 Amd.2-4 – A sample of a Cylindrical Coordinate System. FH is the Frankfurt Horizon as defined in 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 [2].

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

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

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

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

5.9.4 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 defined in this part of ISO/IEC 19794. The values are given in the physical unit mm (millimetre). In the case of Cartesian coordinates ScaleX also has the physical unit mm, in the case of a cylindrical coordinate system ScaleX has the physical unit of rad (degree radian). Each factor is represented by a mandatory four byte float value.

Note that large values of ScaleX, ScaleY or ScaleZ indicate a low resolution in the respective dimension.

Note that 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).

Furthermore note, that 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 Amd.2-4 – 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 Amd.2-5 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 Amd.2-5 – 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 Amd.2-6 – 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 (ms). The field allows the coding of positive as well as negative differences. Here, a negative time difference denotes that the 3D acquisition started before the 2D acquisition. The time difference in milliseconds (ms) is coded in the two's complement system. So, a value of 0x8001 codes the maximum negative temporal difference of −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 Amd.2-7 – The 3D to 2D Image Temporal Synchronicity

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

5.9.10 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 (ms). NOTE It does **not** refer to the synchronicity between the acquisition of the 2D image in the image data block and the 3D data.

The field allows the coding of positive as well as negative differences. Here a negative time difference denotes that the 3D acquisition started before the 2D acquisition. The time difference in milliseconds (ms) is coded in the two's complement system. So, a value of 0x8001 codes the maximum negative temporal difference of −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 Amd.2-8 – The 3D to 2D Texture Temporal Synchronicity

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

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

Table Amd.2-9 – The 3D Acquisition Time

Description	Value
Duration of the 3D acquisition process in milliseconds (ms)	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 ms (milliseconds). A value of 0xFFFF is acceptable and indicates that the field is not specified. NOTE this is **not** the time needed to acquire the 2D image of the image data block.

Table Amd.2-10 – The 2D Texture Acquisition Time

Description	Value
Duration of the 2D acquisition process in milliseconds (ms)	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 Amd.2-11 – 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 Texture Map Spectrum field shall be unspecified.

Table Amd.2-12 – 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: a Range Image, a 3D Point Map, or using 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 Amd.2-13 – 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 contain 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 Amd.2-15 in clause 5.10.7 of this standard. If the existence of an Error Map is specified in the 3D Supplemental Data field, the Vertex Error field shall be present for each vertex.

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

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

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

Table Amd.2-14 – 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 Amd.2-15 – The Error Map

Description	Value
Reserved for future use	0…200
Depth value is interpolated, interpolation type isn't specified	201
Depth value is interpolated, linear interpolation has been used	202
Depth value is interpolated, bicubic interpolation has been used	203
Value of optional Texture Image potentially wrong (texture noisy, overexposure, etc.)	204
Value of optional Texture Image has been corrected by post processing (image processing)	205
Reserved for future use	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 standard 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. To be conformant with this part of ISO/IEC 19794-5 the 2D images of the image data block of any 3D image type have to be conformant with the relevant specifications of the standard 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.

Page 31

Add the following new clauses after Clause 9:

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 this 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 that some of the mandatory fields may still remain unspecified if the appropriate value is set. Please see the specification of the individual fields for details.

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 Amd.2-2.

11.3 Pose of the 3D representation

The rotation of the head in the 3D representation shall be less than +/- 5 degrees from frontal in pitch and yaw (see 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 degrees 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 1 mm.

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

