INTERNATIONAL STANDARD

ISO/IEC 39794-17

First edition
2021-06

Information technology — Extensible biometric data interchange formats —

Part 17:
Gait image sequence data

Reference number
ISO/IEC 39794-17:2021(E)

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.

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

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. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents) or the IEC list of patent declarations received (see patents.iec.ch).

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

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see www.iso.org/iso/foreword.html. In the IEC, see www.iec.ch/understanding-standards.

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

A list of all parts in the ISO/IEC 39794 series can be found on the ISO website.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html and www.iec.ch/national-committees.

The purchase of this ISO/IEC document carries a copyright licence for the purchaser to use ISO/IEC copyright in the schemas in the annexes to this document for the purpose of developing, implementing, installing and using software based on those schemas, subject to ISO/IEC licensing conditions set out in the schemas.

Introduction

Most countries around the world use biometric recognition systems for law enforcement and border control. Many of these systems are not limited to face recognition purposes. To be consistent in such deployments and processes, technical documents, guidelines and best practice recommendations are being developed by different groups. However, these documents are primarily focused on travel documents and related border control systems and the technical and operational issues to be considered when planning and deploying them. Gait recognition is the biometric mode used as a secondary mode in addition to biometric full body recognition or for forensic purposes. Face recognition is the biometric mode best suited to the practicalities of travel documents and automated border processing.

There is little guidance covering the gait imaging for cross-border interoperability or law enforcement services. There is a need for guidance for the use of high-quality digital cameras and video surveillance devices to record gait image sequence data. This document is not restricted to full body gait image sequence data. For example, it can be possible to extract only head movement data for recognition. Gait recognition in this document therefore also covers recognition based on different body parts, e.g. head or limb.

To enable applications on a wide variety of devices, including devices that have limited data storage, and to improve biometric recognition accuracy, this document addresses not only data format, but also scene constraints (lighting, pose, expression, etc.), photographic properties (positioning, camera focus, etc.), and digital image attributes (image resolution, image size, etc.).

A specific biometric profile for cross-border interoperability is required for gait video and still images. Gait image sequence data standardization is required to achieve the threshold quality gait image database records required for automated gait biometric verification and identification. At the moment, border guards record gait video using local practices for gait biometric enrolment, verification and identification.

In order to fully understand the requirements implied in this document it is recommended that the user become acquainted with the following documents: ISO/IEC 39794-16, specifying full body image file formats; ISO 22311, giving information on a common output file format that can be extracted from the video-surveillance contents collection systems to perform necessary processing; the ISO/IEC 30137 series, giving information on the use of biometrics in video surveillance systems; and EN 62676[7] defining video surveillance systems for use in security applications.

This document is intended to provide advice on the use of body image data for gait and upper body movement recognition applications requiring exchange of gait image sequence data and upper body movement data. Typical applications are:

— automated body biometric verification and identification (one-to-one as well as one-to-many comparison),

— support for human biometric verification by comparison of persons based on video and still gait images, and

— support for human examination of video and still gait images with sufficient resolution to allow a human examiner to perform biometric verification.

The structure of the data format is compatible with ISO/IEC 39794-5 and ISO/IEC 39794-16.

This document specifies application-specific profiles including scene constraints, imaging properties and digital image attributes, like image spatial and temporal sampling rates, image size, etc. These modality and application profile specifics are contained in Figures 6 and 7 respectively. Data creation and exchange is described in ISO/IEC 39794-16. The body image data blocks used in encoding gait image sequence data are of type *BodyImageDataBlockType*, which is defined in ISO/IEC 39794-16. This document makes normative reference to other ISO/IEC International Standards.

Information technology — Extensible biometric data interchange formats —

Part 17: Gait image sequence data

1 Scope

This document specifies examples of application-specific requirements, recommendations and best practices in data acquisition applicable to gait image sequence data. Its typical applications include:

a) support for human examination of high-resolution video and still images;

b) support for human biometric verification and identification based on video and still images;

c) automated gait image sequence verification and identification.

This document ensures that image sequences are suitable for human identification and human verification generated by video surveillance and other similar systems.

The following topics are not in scope of this document:

— Definitions for facial and/or full body image related biometric profiles, which are fully covered in ISO/IEC 39794-5 and ISO/IEC 39794-16 respectively.

— Security aspects like digital image sequence electronic signature, Presentation Attack Detection (PAD) and morphing prevention.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 10918-1, *Information technology — Digital compression and coding of continuous-tone still images: Requirements and guidelines*

ISO/IEC 10918-5, *Information technology — Digital compression and coding of continuous-tone still images: JPEG File Interchange Format (JFIF) — Part 5:*

ISO/IEC 14496-1, *Information technology — Coding of audio-visual objects — Part 1: Systems*

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

ISO/IEC 15444-1, *Information technology — JPEG 2000 image coding system — Part 1: Core coding system*

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

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

ISO/IEC 39794-1, *Information technology — Extensible biometric data interchange formats — Part 1: Framework*

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

ISO/IEC 39794-16, *Information technology — Extensible biometric data interchange formats – Part 16: Full body image data*

XML Schema Part 0: Primer Second Edition, W3C Recommendation, October 2004, https://www.w3.org/TR/xmlschema-0/

XML Schema Part 1: Structures Second Edition, W3C Recommendation, 28 October 2004, http://www.w3.org/TR/xmlschema-1/

XML Schema Part 2: Datatypes Second Edition, W3C Recommendation, 28 October 2004, http://www.w3.org/TR/xmlschema-2/

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 39794-1, ISO/IEC 39794-16, and ISO/IEC 2382-37 and the following apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at https://www.electropedia.org/

3.1
3D model
DEPRECATED: 3D image
three-dimensional biometric capture subject body representation that encodes a surface or a volumetric shape in a 3D space

Note 1 to entry: a 3D model can be a heavily processed biometric subject body 3D shape.

3.2
biometric profile
conforming subsets or combinations of base standards used to effect specific biometric functions

Note 1 to entry: Biometric profiles define specific values or conditions from the range of options described in the relevant base standards, with the aim of supporting the interchange of data between applications and the interoperability of systems.

[SOURCE: ISO/IEC 24713-1:2008, 3.9]

3.3
full body recognition
automated recognition of individuals based on their morphology

Note 1 to entry: This can include any or all of the head, torso and limbs.

3.4
gait recognition
automated recognition of individuals based on their manner of walking

3.5
human identification
process of searching through a list of biometric capture subject images to match against an input image(s)

Note 1 to entry: Also known as one-to-many (1: N) searching.

3.6
vignetting
reduction of image brightness or saturation toward the periphery compared to the image centre

4 Abbreviated terms

AVC	advanced video coding
BAP	body animation parameter
BDB	biometric data block
BER	basic encoding rules
CCTV	closed-circuit television
CEN	European Committee for Standardization
CIE	International Commission on Illumination (Commission Internationale de l'Eclairage)
CNN	convolutional neural network
DCI	Digital Cinema Initiatives consortium
DCNN	deep convolutional neural network
DER	distinguished encoding rules
DL	deep learning
EXIF	exchangeable image file format
FAP	face animation parameter
FFM	face features motion
FOV	field of view
GEI	gait energy image
GHM	gesture hand motion
HD	high definition or horizontal deviation angle
HDR	high dynamic range
HMD	head movements dynamic body
HMS	head movements static body
ICS	implementation conformance statement
INTERPOL	International Criminal Police Organization
ISO	International Organization for Standardization
JFIF	JPEG file interchange format
JPEG	image compression standard specified as ISO/IEC 10918

JPEG2000	image compression standard specified as ISO/IEC 15444
JTC	Joint Technical Committee
MP4	ISO/IEC 14496-14 digital multimedia file format used to store video and audio
MPEG	Moving Picture Experts Group
MPEG-4	ISO/IEC 14496-2 video compression format
MTF	modulation transfer function
MTF20	highest spatial frequency where the MTF is 20 % or above
NTSC	National Television System Committee analogue television colour system
PAD	presentation attack detection
PNG	portable network graphics format
RGB	red green blue colour representation
SD	standard-definition television
SFR	spatial frequency response
THz	terahertz
UBM2D	upper body movement in 2D
UHD	ultra-high definition
USAF	US Air Force
VGA	video graphics array image format having width 640 pixels and height 480 pixels
XML	extensible markup language
XSD	XML schema definition

5 Conformance

A BDB conforms to this document if it satisfies all relevant normative requirements related to:

— Its data structure, data values and the relationships between its data elements given in ISO/IEC 39794-16.

— The relationship between its data values and the input biometric data from which the BDB was generated as specified in ISO/IEC 39794-16.

— The application profile-specific conformance specifications given in Clause 8.

A system that produces BDBs is conformant to this document if all BDBs that it outputs conform to this document (as defined above) as claimed in the ICS associated with that system. A system does not need to be capable of producing BDBs that cover all possible aspects of this document, but only those that are claimed to be supported by the system in the ICS.

A system that uses BDBs is conformant to this document if it can read, and use for the purpose intended by that system, all BDBs that conform to this document (as defined above) as claimed in the ICS associated with that system. A system does not need to be capable of using BDBs that cover all possible aspects of this document, but only those that are claimed to be supported by the system in an ICS.

Conformity with this document also requires conformance with the record format specification defined in ISO/IEC 39794-16.

6 Modality specific information

6.1 Purpose

This clause contains modality specific information, where a biometric modality is an information category of a human trait. In general, there are various traits present in humans, which can be used as biometric modalities. There are three human trait categories: the physiological, the behavioural and the combination type of physiological and behavioural modalities. Gait and upper body movement are behavioural modalities.

This clause also describes the requirements and best practice recommendations to be applied for gait and upper body movement image sequence capturing in the application case of enrolment of biometric reference data for feature databases. Conditions for capturing are discussed in more detail in Annex A.

6.2 Practices

The reliable extraction of characteristic features from image sequences and their recognition are important issues in gait and upper body movement recognition. The basic body movement video or a sequence of still images forms the basis for further analysis processing steps. Gait and upper body movement are considered in this document to be the coordinated, cyclic combination of movements that result in human locomotion.

For certain criteria, there may be two different levels: a minimum requirement and a best practice recommendation. The wording is shown in Table 1. The requirement gives the minimum acceptable values or value ranges in order to reach conformance. The best practice recommendation gives values that result in better overall performance or quality, and users are encouraged to adopt best practice values whenever possible.

Table 1 — Summary of wording for minimum requirements and best practice recommendations

Provision	Wording
Requirement	... shall ...
Best Practice	... should ...

6.3 Data models for gait recognition

6.3.1 General

Gait recognition system can be classified depending on the sensors used in three groups, namely; motion imaging (vision)-based, wearable sensor-based and spatial (floor) sensor-based. The motion imaging (vision) can be divided into two groups, namely: appearance-based methods and model-based methods. The appearance-based method can be also subdivided into two types; state space methods and spatiotemporal methods[9]. As stated in the Scope, this document is restricted to the motion imaging-based gait recognition, which may use the whole available electro-magnetic spectrum available, not only the visible bandwidth. The scope of this document is marked with bold text and continuous box outline in Figure 1.

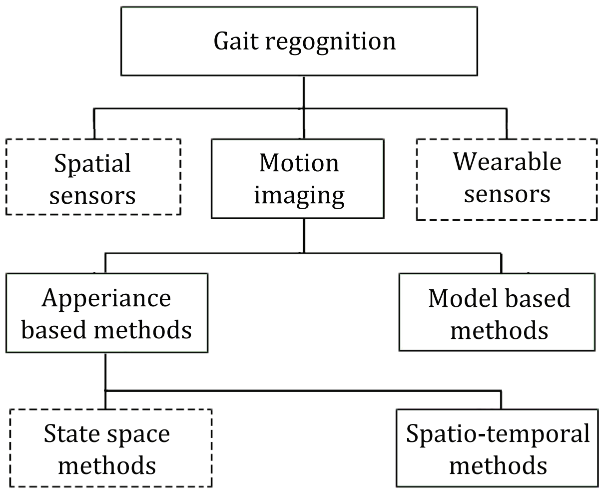


Figure 1 — Classification of gait recognition systems.

6.3.2 Model-based methods

Model-based approaches build a human body gait model and the extracted features of gait sequences are fitted to that model. These methods are not sensitive to the individual's appearance and clothing but have high computational cost. It is hoped that the use of machine learning will enhance both the creation of models and the least error model selection.

Model-based feature extraction is used to extract human joints (vertex positions). A vision-based system for human motion analysis consists of three main phases: detection, tracking and perception. In the last phase, a high-level description is produced based on the features extracted during the previous phases from the temporal video stream. Marker-based solutions rely primarily on markers or sensors attached at key locations of the human body.

Gait image sequence enrolment and identification using visual surveillance require the deployment of an automated marker less vision system to extract the joints' trajectories. Automated extraction of the joints' positions is a difficult task as non-rigid human motion encompasses a wide range of possible motion transformations due to its highly flexible structure and to self-occlusion. Clothing type, segmentation errors and different viewpoints pose a challenge for accurate joint localization. For a model-based approach, a shape model is *a priori* established to match real images to this predefined model, and thereby extract the corresponding features once the best match is obtained[10].

6.3.3 Appearance-based methods

Appearance-based methods or model-free gait recognition methods work directly on the gait sequences. They do not use a model for the human body to rebuild human walking steps. These methods have the advantage of low computational cost in comparison with model-based approaches, but the disadvantages are sensitivity to changes in clothing and appearance. Applying an averaged silhouette of a biometric subject during a gait cycle or using information obtained from a submillimetre image enhances the silhouette image accuracy.

The decision to omit the state space from the scope of this document is based on the present status of non-conformance regarding the use of state space results. Various linear combinations of a system's state variables can be used to span its state space and different reconstruction methods can yield different solutions[11], rendering their comparison a challenge. There should be consensus on how to reconstruct the state space for gait dynamics in order to standardize state space methods.

6.4 Data flow of gait recognition

Figure 2 illustrates the components and data flow between the components in a biometric gait image sequence processing system.

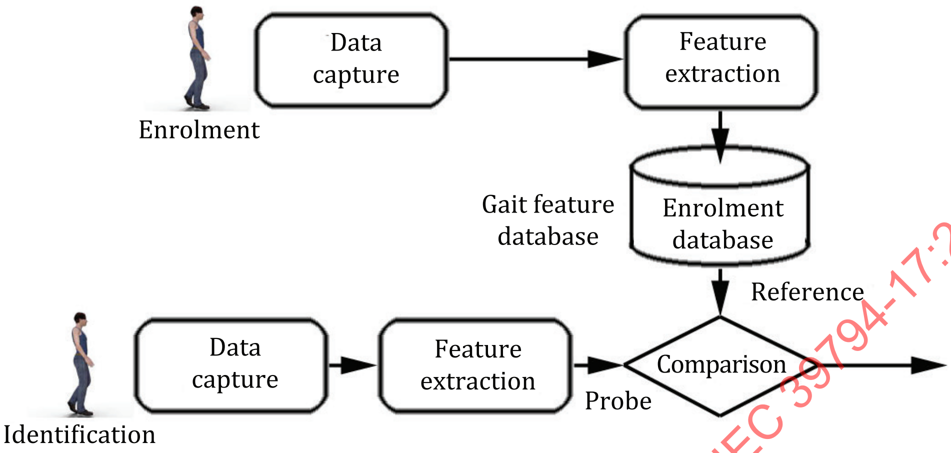


Figure 2 — Components of a gait image sequence biometric system.

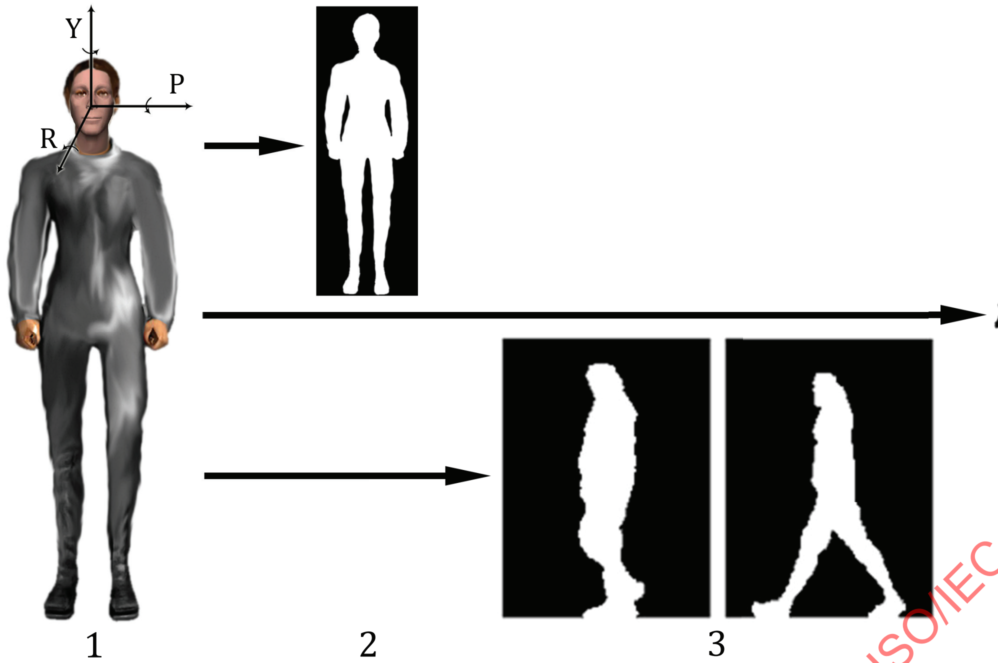
Comparison methods may use conventional feature-based template sets or deep convolution neural network (DCNN) feature vectors. After the feature vectors are generated using gait signatures and DCNN processing then the comparison is based on one of the many basic machine learning classification algorithms e.g. Bayesian classifier or Euclidean classifier. See A.2, Deep Convolution Neural Network (DCNN) presentations.

6.5 Body tree concept for gait

Gait imaging systems utilize 2D recordings or 3D models for human examination and for automated gait verification and identification. Instead of using representations as isolated entities a more organized way is to utilize the body tree structure.

For example, multimodal biometric human verification or identification may use face features, full body features, full body gait and head movement. The results should be fused at various levels of fusion, such as comparison score level, feature level and decision level. Submillimetre imaging should be used to address the problem of clothing variation effect on gait matching.

Figure 3 illustrates the possibilities offered by full body images and videos, which provide a wide selection of biometric features for various gait-related processes.



Key

P pitch around the side-to-side x axis

Y yaw around the vertical y axis

R roll around the front-to-back z axis

1 pose

2 appearance

3 gait

4 structure

Figure 3 — Full body features for various processes

Standard poses, element structures and data formats help the parsing of the body tree data into body part representations and landmarks. Parsing can be achieved using methods utilizing algorithms which process the human body as an assembly of parts. Segmentation can be used as a pre-processing step.

Both static full body and dynamic gait cues of body biometrics may be independently used for recognition. Fusion of static and dynamic body biometrics for gait recognition can give better results if the combination strategy is carefully balanced and the score-summation-based rule is used, for example[14].

6.6 Camera image sequence requirements

The original camera image sequence is saved whenever possible without any additional cropping, rotation or other image processing. The full body pose shall be between 60 % and 95 % of the vertical length of the image during enrolment. The whole-body height and width shall be visible. For video recordings, both portrait and landscape camera orientation are acceptable.

The set of photographs shall include at least one recording of the subject in a standard walking pose: (frontal full profile, left full profile, right full profile, back full profile). Additionally, a submillimetre wavelength recording may be included.

Gait recognition, upper body movement recognition and full body recognition can be paired to form a multi-mode biometric process in order to improve the performance of a biometric system. If the person's facial area is not visible or the number of pixels in a video surveillance or other security camera still image is too low, then body silhouette can be used for identification or verification purposes.

Meeting the requirements set for any camera system requires measurements to be taken and analysed. Image sequence acquisition measurements are described in Annex C.

6.7 Gait recognition recordings

6.7.1 General

A gait recognition silhouette is the image of a person represented as a solid shape of a single colour, usually black. The edges of a silhouette match the outline of the subject. Gait recognition walk-through video recording is recommended to improve the performance of both gait recognition and full body photometric recognition.

6.7.2 Gait and upper body movement image encoding

There are several image encodings which shall be used instead of non-standard formats e.g. bitmaps defined in an ad-hoc way or ambiguous formats e.g. TIFF:

a) The JPEG image sequence in Sequential baseline (in accordance with ISO/IEC 10918-1) mode of operation and encoded in the JFIF file format (in accordance with ISO/IEC 10918-5);

b) The JPEG-2000 image sequence in Part-1 Code Stream Format (in accordance with ISO/IEC 15444-1), lossy or lossless, and encoded in the JP2 file format (the JPEG2000 file format);

c) The PNG image sequence in Portable Network Graphics format (in accordance with ISO/IEC 15948), lossless, and encoded according to the Portable Network Graphics (PNG) Functional specification;

d) The MPEG-4 video in AVC/H.264, in accordance with ISO/IEC 14496-10 defined format;

e) The MP-4 video in accordance with ISO/IEC 14496-14 defined format; and

Gait Recognition Landmark Points should be determined on images before compression is applied. Landmark Points should be included in the record format if they have been accurately determined, thereby providing the option that these parameters do not have to be re-determined when the image is processed for body recognition tasks. The Landmark Points should be determined by computer-automated detection mechanisms followed by human validation when necessitated by the legal requirements. At the moment, there are no single recommendations for the gait recognition landmark points.

6.7.3 Gait and upper body camera image resolutions

The most frequently used frame rate in digital video recording is 25 frames per second. Pixel aspect ratio is normally 1:1. However, in several video standards the pixel is defined as non-square. For example, a pixel aspect ratio of 0,90 is used for NTSC to display a frame size of 720 pixels x 480 pixels (DV) or 720 pixels x 486 pixels (D1) for 720 pixels x 540 pixels displayed in 4:3 format. Most digital still image cameras are able to record video. Submillimetre (THz) cameras and scanners have limited image sizes in pixels due to the terahertz wavelength resolution constraints. THz frames are typically DV size.

Image orientation is generally not a problem as JPEG EXIF metadata show the camera orientation. MPEG-4 AVC/H.264 (ISO/IEC 14496-10) implementations for video coding allow frame extraction for biometric sample comparison processing to take place. MPEG-4 Part 14 or MP4 is a digital multimedia format most commonly used to store video and audio. MPEG-4 Part 14 (formally ISO/IEC 14496-14) is a standard specified as a part of MPEG-4, in accordance with ISO/IEC 14496-1 and ISO/IEC 14496-2. MP4 is the related file format.

Table 2 shows the most common digital video formats and respective resolution, aspect ratio and pixel size information. Figure 4 shows the VGA, HD and 4K frames placed on a single 8K frame.

Table 2 — Comparison of digital video recording formats

Video format name	Resolution (pixel)	Display aspect ratio	pixels
VGA resolution	640 × 480	1.33:1 (4:3)	307 200
HD 720 p	1280 × 720	1.78:1 (16:9)	921 600
HD 1080 p	1920 × 1080	1.78:1 (16:9)	2 073 600
DCI 2K	2048 × 1080	1.90:1 (19:10)	2 211 840
UHD 4K (UHD-1)	3840 × 2160	1,78:1 (16:9)	8 294 400
DCI 4K	4096 × 2160	1,90:1 (19:10)	8 847 360
UHD 8K (UHD-2)	7680 × 4320	1,78:1 (16:9)	33 177 600

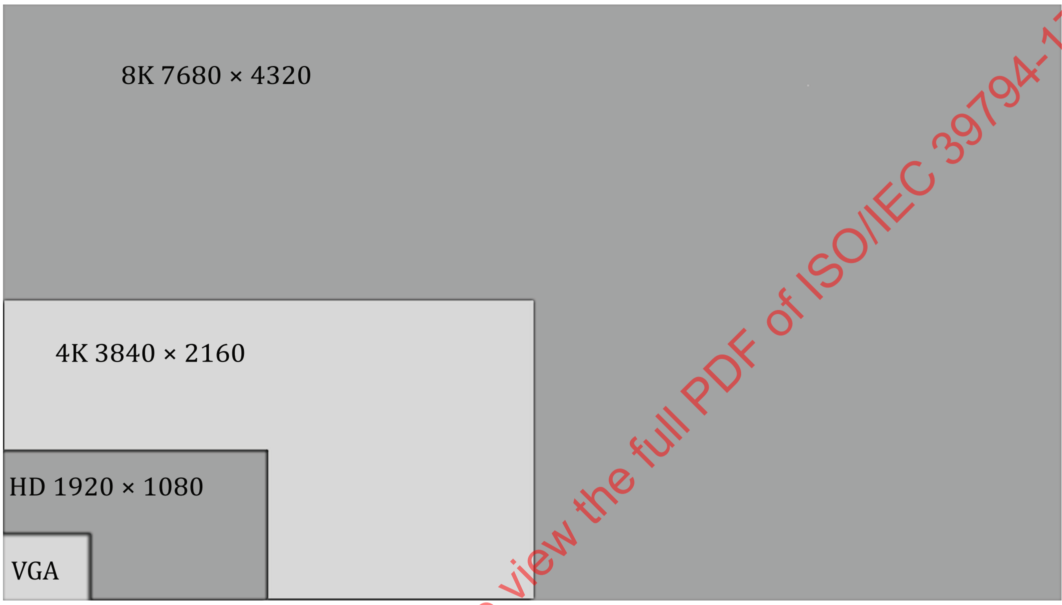


Figure 4 — Comparison of digital video frame sizes

6.8 Gait modality

6.8.1 General

Gait recognition and full body recognition can be paired to form a multi-mode biometric process in order to improve the performance of a biometric system. If the person's facial area is not visible or the number of pixels in a video surveillance or other security camera still image is too low, then body silhouette can be used for identification or verification purposes.

High resolution level-50 and 4K still images are suitable even for facial biometric recognition processes. In automated processing, it can be necessary for the parsing of the body tree to occur first, which in practice involves the detection of all body parts and forming a body tree model. In this case the facial and upper torso is used to get the facial image. Some face recognition programs include this feature.

For example, in order to compare two video frames, including one reference video stream of a person seen in various full body poses and one CCTV video probe showing only the upper torso and head, it would be the right decision to apply head movement analysis on both video streams.

6.8.2 Gait silhouette

A gait recognition silhouette is the image of a person represented as a solid shape of a single colour, usually black. The edges of a silhouette match the outline of the subject. Gait recognition walk-through video recording is recommended to improve the performance of both gait recognition and full body photometric recognition.

To ensure that the gait sequence captures the body movement in detail, it is recommended that the sequence be captured at the rate of 30 frames per second. This is a typical frame rate used in gait research databases such as the CMU MoBo database[15] and USF HumanID gait database[16].

In order to capture all the details of a gait signature, a minimum of one full gait cycle, i.e. two full steps, shall be captured. Figure 5 illustrates the silhouettes of the phases of one full gait cycle showing from left to right stance phases (1, 3 and 5) and between those, swing phases (2 and 4).

Key

1 stance phase

2 swing stage

3 stance phase

4 swing stage

5 stance phase

Figure 5 — Illustration of the phases of a full gait cycle

Various automated methods have been developed for biometric gait recognition. Some methods use the image data as an input while others only use silhouettes. Also, automated methods can use either aligned images or non-aligned images. The capture process should allow for any method to be used for automated recognition, therefore the gait sequence should be captured with a stationary camera.

The side view is the most discriminative[17] view of a gait sequence. The subject should be captured at least using side-view. The subject should be instructed to walk on a straight line perpendicular to the camera line of sight as illustrated in Figure 6. When a tread mill is used for walking it is easier to maintain a stationary view of the person in the middle of the frame. As an option, make a similar video recording showing frontal and back (dorsal) views in addition to the side (lateral) view.

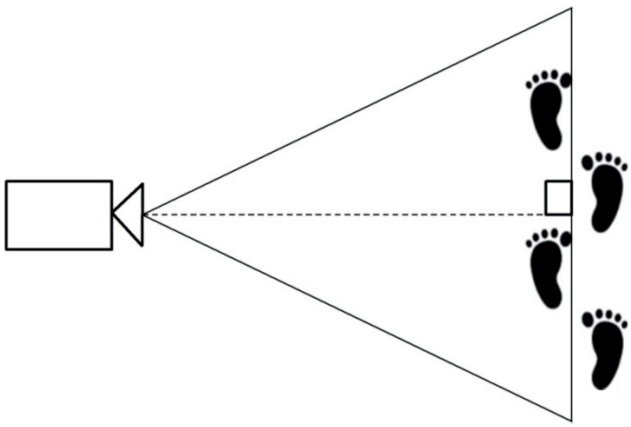


Figure 6 — Top view of the camera and the person's walking path

By combining appearance and motion in a spatiotemporal way it is possible to achieve better results than using a single modality in the most difficult scenarios, where there is variation in both appearance and dynamics[12], [14].

Walk-through video analysis is less time-consuming if illumination and background variations are minimized. There are methods to reduce the effects of illumination variations and dynamic backgrounds[13] in video surveillance video material. Background subtraction is a challenging task, especially in complex dynamic scenes that can contain a moving background, vegetation, rippling water, etc. It is recommended to make studio quality recordings for enrolment so that it is not necessary to pre-process the saved video in order to correct these problems.

6.8.3 Surveillance systems

In video surveillance systems, multiple camera operation is essential for multiple capture zones. This may include ‘layers’ of cameras along the path a target subject is expected to take to allow for multiple detection opportunities (e.g. as the target subject walks along a corridor towards the cameras). This is useful not only if there are issues with frame rate and dropped frames, but also to track a target subject if an alert is triggered, especially if there is significant detection latency.

Multiple cameras are also used for any single video surveillance capture zone that is too wide for one camera to provide sufficient resolution of the face for the required performance levels, and to try and compensate for target subjects that may be facing different directions when traversing a particular camera’s field of view or depth of field. Such target subjects may be deliberately trying to avoid the cameras or they may simply be unaware of their presence.

Gait analysis is deployed for people identification in multi-camera surveillance scenarios. View-point independent rectification is used to calculate side view coordinates for multi-camera video frames, for example. Low frame-rate (1-5 fps) video recordings made with still image cameras or video surveillance systems are compatible with normalized gait silhouette sequences.

Typical recording time for enrolment is in the order of 6 seconds. Identification rate improvement after 5 fps approaches that of the higher frame rate videos[18]. Low frame-rate high resolution level-50 and 4K video storage size is much smaller than full frame rate video for the same recording time thus making the low frame rate more economic for data base storage.

SD resolution (640 x 480 px) or higher can be used for gait sequence enrolment recording. Low frame-rate gait sequence can be constructed by down sampling the 30 fps enrolment database sequence at a specified interval. This way, it is possible to use low frame-rate video data from video surveillance camera as a probe, without using temporal reconstruction.

6.9 Upper body movement modality

6.9.1 General

Biometrics identification using upper body dynamics concentrates on the head area and face features motion as shown in Figure 7. Typical dynamic face features are blinking, eye movements and gazing. Head movements are either static, while the person is not walking, or dynamic when walking, as a result of gait. Movement of hands can also be disordered or dynamic free movement or more controlled gestures.

Figure 7 — Upper body movement: key areas and landmarks

6.9.2 Face Features Motion (FFM)

In a human face we can identify many unique objects such as eyes, nose, lips. The combination of the sizes of these objects and the distance between them is the base of face recognition. The dynamics of these objects between themselves is unique. FMMs are typically blinking of the eyes, eye movement, gaze changes and movements of the person's lips. In some cases, for example when the person talks or smiles, FFM could be the movement of scars and skin features such as wrinkles and other marks on the skin.

6.9.3 Head movement recognition

Head movement recognition is divided into two separate parts:

— Typical micro head movements of the person. A typical micro head movement can be noticed in all body poses, when standing still, moving or sitting. Each person has unique neck and head movements, i.e. head poses that are based on the movement of the neck in the pitch, roll and yaw axes of the head.

— Head and body movement is a result of the gait, when the person is walking. The body moves when the person is their moving legs, and as the head is attached to the rest of the body, it moves along with the body. Each person has unique head movements which are based on their unique gait. These movements can be categorized through the location changes of the head in up-down and left-right direction in the overall frame of the video.

6.9.4 Head Movements Static Body (HMS)

Each person has unique micro neck and head movements, or head poses, that can be valuable when a person is not walking or when the body is not otherwise moving. HMS poses are based on the movement of the neck in the pitch, roll and yaw axes of the head and the natural reflex of the muscles. Dynamic DL and optical flow are used to obtain the uniqueness of the micro head motion relative to the static body in the HMS scenario. The first step needs to be the identification of an accurate head pose.

6.9.5 Head Movements Dynamic Body (HMD)

Each person has unique head movements created as a result of the specific gait from the movement of the walking body. Dynamic DL and optical flow are used to obtain the uniqueness of the head motion when the person is walking. One of the challenges in identifying the dynamic body head movement (HMD) is the need to dynamically subtract the static body head movements (HMS), which are typical to the person, from the total head movement when a person is walking and be left with the clean HMD as result of gait. This clean HMD can also be used to identify the person.

The extracted HMS itself is an additional identifying factor.

HMS and HMD can usually be well separated as they are usually in very different mechanical frequency ranges and at very different amplitudes.

6.9.6 Hands movement

Each person has a unique way of moving their hands, especially when walking and also when standing still. Dynamic DL and optical flow are used to obtain the uniqueness of the movement of the hands. In many cases this hand movement can also be used as gesture hand motions (GHM) to identify specific gestures.

7 Profile-specific information

7.1 Purpose

7.1.1 General

This clause contains an overview of the gait and upper body movement recording profiles. For an explanation of the data elements see ISO/IEC 39794-16. Data conventions follow the definitions given in ISO/IEC 39794-16.

7.1.2 Gait representations

Each gait image sequence data block shall pertain to a single biometric subject and shall contain one or more 2D image sequences of a human body and/or human-body related information. Thus, each BDB can contain zero or more geometric representations (range image, point maps, vertex representations, neural network representations).

Samples can contain post-processed data as video recording devices have pattern recognition, human classification, and other processing features embedded in the camera software. In such cases, the MPEG-4 samples may contain Face Animation Parameters (FAPs) and/or Body Animation Parameters (BAPs). When further processing is done using neural network in the camera software, the end result after the processing can be a neural network feature vector representation. In that case the Gait Neural Network Feature Vector Representation Profile shall be used. For these cases, all necessary data elements are described in ISO/IEC 39794-16.

Using convolutional neural network-based approaches it is possible to extract and aggregate useful information from gait silhouette sequence images instead of representing the gait process by averaging silhouette images[26]. The available biometric sample format is used to decide which profile shall be applied.

7.1.3 Scene requirements

The specification of a certain background is not normative for the creation of gait image sequences. A consideration of the background is important for computer-based gait recognition because the first step in the computer gait recognition process is the segmentation of the body from the background.

Lighting should be equally distributed on the body to allow precise silhouette processing. There should be no significant direction of the light from the point of view of the camera. A silhouette is the image of a scene representing as a solid shape of a single biometric subject, usually in black, with its edges matching the outline of the subject.

7.2 2D gait image sequence profile

7.2.1 General

The *Gait2D* profile is used in cases where the camera produces sequential images. It can be necessary to process the data in order to obtain necessary body animation points or to compress data into face or body landmark point format. When only face or body animation parameters are needed, then the post-processed gait video and/or still images contain a set of landmark point coordinates or similar information for the reconstruction of facial or bodily movement.

The ISO/IEC 39794-17 (this document) BDB follows the ISO/IEC 39794-16 BDB record format definitions.

7.2.2 Gait image sequence representation profile requirements

Rather than impose a particular hardware and lighting capture system, this subclause specifies the desired output image or image sequence properties. The requirements and recommendations apply to digital images when camera system output is in image format.

Neural network processed video and/or sequence of still images is represented either in reconstructed video and/or sequence of still images or in a neural network feature vector format. If the sample is represented in one of the above-mentioned image formats, then the requirements of this subclause shall be applied.

Modality specific information presented in Clause 6 applies to this profile.

7.2.3 Post-acquisition processing

Basic recording is usually post-processed by removing the background and silhouetting the full body image on each video frame. This can be achieved by the camera firmware using conventional methods or neural networks. Local processing using neural networks is always possible using local feature vector or tensor formats. However, gait image sequence output data for information interchange using ASN.1 or XML shall follow the format specifications given in ISO/IEC 39794-16.

While classical methods typically derive gait signatures from sequences of binary silhouettes, it is possible to use convolutional neural networks (CNN) for learning high-level descriptors from low-level motion features, i.e. optical flow components. Some experimental results indicate that using spatiotemporal cuboids of optical flow as input data for CNN allows state-of-the-art results to be obtained on the gait task with an image resolution eight times lower than low resolution VGA frames, i.e. 80×60 pixels[22]. The two-stream architecture for video classification[25] is shown in Figure 8.

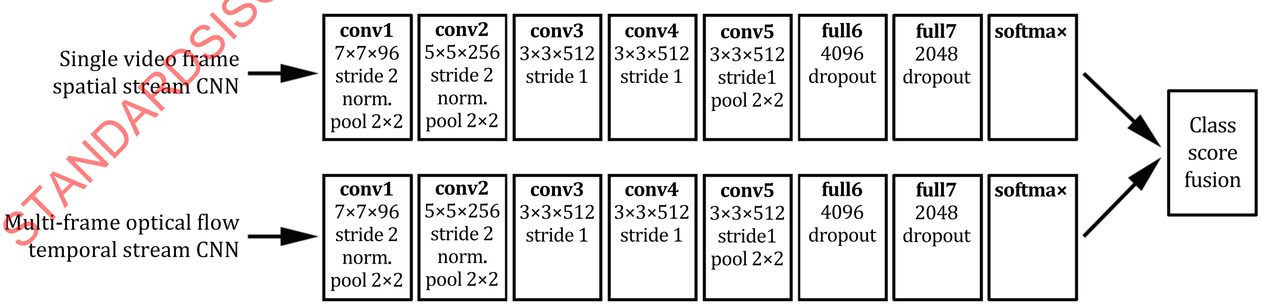


Figure 8 — Two-stream architecture for video classification[25].

Gait based human identification via similarity learning by deep convolutional neural networks (CNNs) using a OU-ISIR dataset has resulted in recognition rates of over 90 %[23]. OU-ISIR is currently one of the largest datasets available in the literature for gait recognition, with 4 007 subjects. On this dataset, the average accuracy of the used method under identical view conditions is above 98 %, and the one for cross-view scenarios is above 91 %. The method described also has promising results on the University of South Florida (USF) gait dataset, containing gait sequences imaged in a real outdoor scene.

In order to use biometric feature vectors for recognition it is important to use the same CNN architecture for references and probes. For example, if the architecture used in creating the reference vectors in the data base is based on a certain set of convolution layers and fully connected layers then the probe vector shall be created the same way. Without standardized architectures, the use of CNN for gait recognition is limited to local databases, all of which are using the same architecture.

7.2.4 Neural network training and testing

Some databases that are widely used are University of South Florida (USF) Gait Dataset, Chinese Academy of Sciences (CASIA) Gait Dataset, Osaka University (OU) and the Institute of Scientific and Industrial Research (ISIR) joint OU-ISIR Gait Database, and Southampton University (SOTON) Gait Dataset. These datasets can be used for neural network training and testing. With a group of labelled multi-view human walking videos, it is possible to train neural networks to recognize discriminative gait patterns used to identify subjects. Neural networks are able to automatically learn to recognize the most discriminative changes of gait features to predict similarity given a pair of them.

The CASIA Gait dataset has four class datasets: Dataset A, Dataset B (multi-view dataset), Dataset C (infrared dataset), and Dataset D (foot pressure measurement dataset). CASIA Dataset A (former NLPR Gait Database) was created on December 10th 2001, including 20 persons. Each person has 12 image sequences, 4 sequences for each of the three directions, i.e. parallel, 45° and 90° to the image plane. Dataset B is a large multi-view gait database, which was created in January 2005. There are 124 subjects, and the gait data was captured from 11 views. Three variations, namely view angle, clothing and carrying condition changes, are separately considered. Besides the video files, the dataset provides human silhouettes extracted from video files. Dataset C was collected by an infrared (thermal) camera in July-August 2005. It contains 153 subjects and takes into account four walking conditions: normal walking, slow walking, fast walking, and normal walking with a bag. The infrared videos were all captured at night. Dataset D was collected synchronously by camera and foot scan in July-August 2009. It contains 88 subjects and takes into account real surveillance scenes and wide age distribution. In Figure 9 the CASIA dataset GEI silhouette example row of images is showing a subject walking from different viewing angles. The gait energy image representation is formed by averaging the size-normalized and centre aligned silhouettes of a gait cycle.

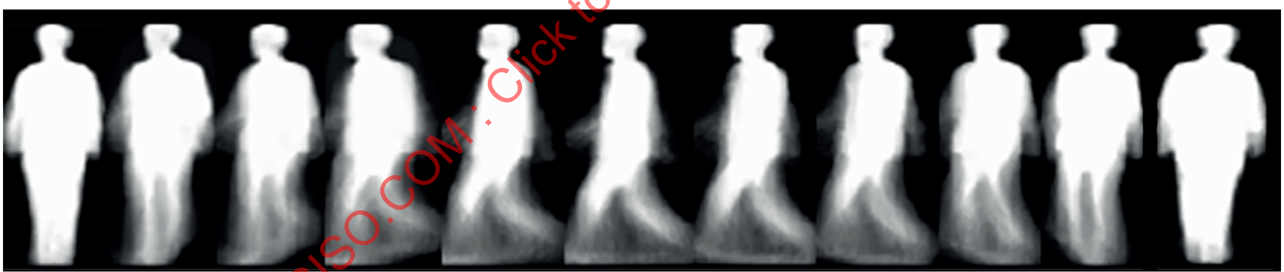


Figure 9 — CASIA GEI dataset image example

The OU-ISIR Gait Database contains a dataset consisting of 10 307 subjects (5 114 males and 5 193 females of various ages, ranging from 2 to 87 years) from 14 view angles, ranging from 0°-90° and 180°-270°. Gait images of 1280 × 980 pixels at 25 fps are captured by seven network cameras (Cam1-7) placed at intervals of 15° azimuth angles along a quarter of a circle whose centre coincides with the centre of the walking course. Its radius is approximately 8 m and height is approximately 5 m. The entire OU-ISIR Gait dataset is divided into two disjoint subsets, i.e. training and testing set, that both have almost the same size.

The University of Southampton 3D gait dataset uses eight synchronized cameras to capture gait and additional cameras to capture images from the face looking towards the subject at the end of the tunnel and one ear with a camera at one side of the tunnel, as an individual walks through the multi-biometric tunnel. Whilst a subject is inside the tunnel their gait is recorded by eight cameras, allowing the reconstruction of 3D volumetric data. The gait cameras all have a resolution of 640 × 480 pixel and capture at a rate of 30 frames per second. they are connected together over a network employing synchronization units to ensure accurate timing between cameras. Video is also captured of the subject's face and upper body using a high resolution 1600 × 1200 pixel camera, enabling face recognition. A 1600 × 1200 pixel snapshot is taken of the subject's side of the head, which can be used for ear biometrics.

For exact timing of the events, Capture date/time block abstract values should indicate the date and time when the capture of the representation was completed in UTC (see the ISO 8601 series) to a precision of up to one millisecond.

7.3 UBM 2D upper body movement profile

The Upper Body Movement (UBM2D) representations contain the 2D Representation Data, the 2D Capture Device Technology and the 2D Image Information. The 2D Representation Data, the 2D Capture Device Technology and the 2D Image Information elements are described in 2D Gait Representation.

Upper body movement subclasses are

— face features motion (FFM)

— head movements static body (HMS)

— head movements dynamic body (HMD)

— gesture hand motion (GHM)

8 Encoding

8.1 Tagged binary encoding

ISO/IEC 39794-16:2021, A.1 specifies an ASN.1 schema in which the abstract data elements are constrained by ASN.1 types defined within one of the following standards: the ASN.1 standard ISO/IEC 8824-1, ISO/IEC 39794-1, ISO/IEC 39794-5, ISO/IEC 39794-16, or by this document.

The tagged binary encoding of full body image and gait image sequence data shall be obtained by applying the ASN.1 Distinguished Encoding Rules (DER) defined within ISO/IEC 8825-1 to a value of the type BodyImageDataBlock defined in the given ASN.1 module. The DER encoding of each data object has three parts: tag octets that identify the data object, length octets that give the number of subsequent value octets, and the value octets.

According to ISO/IEC 8824-1, each ASN.1 module begins with a module name that identifies the module for human beings. ASN.1 type names are in upper camel-case notation derived from subclause titles in the abstract data elements clause. If a name starts with a number, then this component is put to the end of the name. Names of components of a structured type are in lower camel-case notation and are unique within the scope of this structured type.

The ASN.1 modules defined in ISO/IEC 39794-16 allow the extension of definitions in a backward and forward compatible way as specified in ISO/IEC 8824-1. The new edition of ISO/IEC 39794-16 may specify new additional data elements and values and deprecate the use of previously defined optional data elements and values. ASN.1 modules in the new edition of ISO/IEC 39794-16 shall include all data elements and values specified in the previous edition.

The ASN.1 module can be retrieved from https://standards.iso.org/iso-iec/39794/-16/ed-1/en/.

Encoding examples are contained in Annex B.

8.2 XML encoding

An XML document encoding body data shall obey the XSD given in ISO/IEC 39794-16:2021, A.2. Annex B.2 examples use XSD schemas, in which the abstract data elements are constrained by XML types defined within one of the following standards: W3C Recommendations, XML Schema Parts 1 and 2, ISO/IEC 39794-1, ISO/IEC 39794-5, ISO/IEC 39794-16, or this document.

Binary data shall only be encoded as base 64 and stored as a text string in an xs:element, which itself has the underlying type of ‘xs:base64Binary‘.

Example <xs:element name="data" type="xs:base64Binary"/>

For avoidance of doubt, other methods of encoding binary data such as xs:hexBinary or proprietary extensions which support binary data encoding (i.e. XOP) are not permitted.

The XML data type “base64Binary” is not supported in SPARQL (SPARQL Protocol and RDF Query Language). Recommendation ITU-T X.694 and identical ISO/IEC 8825-5 define rules for mapping an XSD Schema (a schema conforming to the W3C XML Schema specification) to an ASN.1 schema in order to use ASN.1 encoding rules such as the Basic Encoding Rules (BER), the Distinguished Encoding Rules (DER), the Packed Encoding Rules (PER) or the XML Encoding Rules (XER) for the transfer of information defined by the XSD Schema.

The XSD module can be retrieved from https://standards.iso.org/iso-iec/39794/-16/ed-1/en/.

Encoding examples are contained in Annex B.

9 Registered BDB format identifiers

The registrations are listed in ISO/IEC 39794-16 to identify the full body image data interchange formats referenced in this document.

For details, see ISO/IEC 39794-16.

Annex A
(informative)

Conditions for capturing

A.1 Recommendations for Gait Recording Systems

A.1.1 Architecture

The aspects that can be considered by the definition of full body photography system architecture described in ISO/IEC 39794-16 also apply to gait recording. These are:

— biometric capture sub-process, carried out by a video capture unit,

— image formatting sub-process, carried out by a video application program,

— visualization of process and results, both for operator and the subject of biometric authentication process,

— integrity of video system,

— denial of service (system is supposed to be used only by authorized operator), what can be done, for example, with biometric verification of operator,

— connection to other systems (databases, etc.) with the focus on protocols and the integrity of such connection.

Only the first two aspects are included within the scope of this document. The other aspects are covered by the documents in the Bibliography and are not considered in detail in this document.

A.1.2 Usability and accessibility

Accessible systems should be designed to be equitable in use for subjects who have permanent or temporary physical or psychological inabilities. They should be simple to use and with a wide tolerance of operation. For subjects that cannot use the biometric system, alternative systems are necessary and should be provided.

General guidance on these aspects is given in ISO/IEC TR 24714-1. Pictogram recommendations can be found in the ISO/IEC 24779 series.

A.1.3 Practical applications

A.1.3.1 General

In practice, full body images and walk-through videos are taken from crime suspects and saved to databases. The biometric capture process may involve a single biometric capture device (camera) or several devices. Biometric features are extracted from high resolution samples. High resolution images and 4K videos carry enough information for various biometric processes. Biometric reference is formed as the stored biometric samples are attributed to a biometric data subject and used as the object of biometric comparison.

A.1.3.2 Body vs. gait

Gait image sequences represent behavioural walking features, which are easier to extract from a low-resolution recording than full body biometric recognition features in a single frame. A good quality full body image is capable of showing even facial features for face recognition purposes. At the moment, video surveillance systems rarely produce sharp facial images. Good quality full body images are easier to extract from HD video frames.

Body images appear in video surveillance videos more often than clear face images. Faces and body images are added to the watch lists from observation of behaviours in the video material. In order to increase the international interoperability of the biometric samples, it was necessary to create a standard describing the biometric sample formats for full body images and video recordings. It is good to note that video is always created by taking still images in certain intervals defined by the frame rate. The difference between still images and video frames is mainly the time between individual images. Setting the timeline right, it is possible to compare still images and video frames for identification or verification purposes. Time values should indicate the date and time when the capture of the representation was completed in UTC (see the ISO 8601 series) to a precision of up to one millisecond.

Depending on the cultural requirements, persons are photographed wearing clothing or near naked. According to INTERPOL guidelines[29], photographic and video recording of bodies at a disaster site and within a mortuary is important both for evidence and because it can help to establish the cause of an incident. Gait image sequences do not necessarily require high quality recording.

State-of-the-art biometric recognition algorithms can be used to select images of people with almost no useful identifying information in the face. In these cases, recognition of the face alone is by chance, but recognition of the person based on larger part of the body is accurate, according to the research work carried out at the University of Texas in Dallas[30]. The assumption in the automatic face recognition community has been that all important identity information is in the face. These results should point us toward new ways to improve automatic recognition.

SD video frame comparison with high quality reference images gives better results than comparing two SD level images. At the moment, 4K DSLRs capable of taking high resolution still images and 4K video are available at a reasonable price. Video surveillance systems are set to evolve from HD to 4K in the near future and therefore the resolution level of still images should be high. A passport quality face image is obtained by cropping the 4K full body frontal image of a person.

Data formats for biometrics in video surveillance systems (ISO/IEC 30137-3) and high efficiency coding (ISO/IEC 23008-12) are standardized for data format integration work.

See ISO/IEC 39794-16 for gait video imaging technical requirements, which are the same as full body photography requirements for enrolment.

A.1.3.3 Recognition methods

In silhouette-based recognition, camera data are post-processed to obtain the sequence needed and to disentangle the identity-unrelated factors which alter gait appearances. These factors can be grouped into:

— biometric subject-related ones, such as walking speed, dressing and carrying conditions;

— device-related ones such as different frame rates and imaging resolutions; and

— environment-related ones such as illumination conditions and camera viewpoints.

In silhouette cases, the data is saved as a post-processed gait video recording containing a set of images as described in this document.

Model based recognition body model representations shall use only those data types which are specified in this document. It is strongly recommended to convert any other manufacturer-specific data to conform with one of the standard data types given in this document.

Upper body movement recognition is post-processed using the methods described in this document.

A.1.4 Submillimetre imaging

Terahertz (THz) radiation known as submillimetre radiation consists of electromagnetic waves within the ITU-designated band of frequencies from 0,3 to 3 terahertz (300-3000 GHz). Wavelengths of radiation in the terahertz band correspondingly range from 1 mm to 0,1 mm and this wavelength also sets the theoretical resolution limit for the images. Terahertz radiation is located in the terahertz gap between microwaves and infrared light waves in the electromagnetic spectrum.

Imagers in this document are divided in two main categories: active and passive systems. The active imagers are based on a radar configuration, with synthetic or optical imaging schemes, whereas passive ones operate as cameras. The main challenge in all these systems is to achieve a large field of view (FOV) with satisfactory spatial resolution and fast image acquisition. In Figure A.1, a submillimetre image shows suspicious material hidden under subject's clothing in the middle.

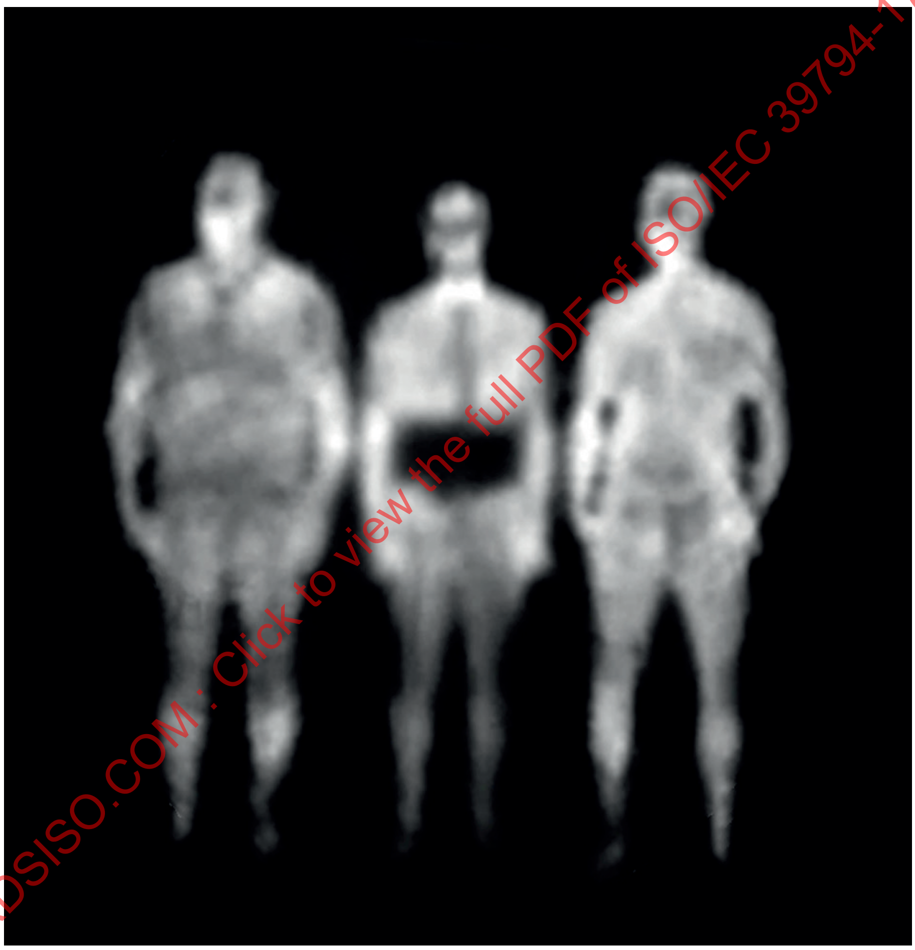


Figure A.1 — Submillimetre image of three subjects

The THz beam can traverse non-metallic materials like clothing fabrics, paper and cardboard, plastic and ceramics. It cannot penetrate liquid water or metal. The imaging area depth is typically less than 10 m. At a distance of 5 m it is possible to take a submillimetre picture of 2 m height using a passive system.

This document focuses on passive submillimetre/THz imaging system requirements that are shaped and driven by the imaging phenomenology. While many passive submillimetre imaging systems have been constructed throughout the world that can produce high-quality, reasonable-contrast and low-noise imagery given adequate frame times of seconds to minutes, few passive systems are capable of producing useful imagery at real-time video rates. The following paragraphs provide a short description of the critical receiver sensitivity levels required to provide useful, real-time video rate passive imagery in the scenarios explored. The measurements were taken at the U.S. Army Research Laboratory[21].

Measurements have revealed that clothing is adequately transparent at frequencies up to approximately 300 GHz (based on a 3-dB criterion). Microphotographs reveal roughness on the scale comparable to a THz wavelength, particularly in the coarse filamentary pattern of nylon and the interlaced weave of rayon and silk. As these sample types were found to be the most transparent, it is surmized that any surface scattering must be fairly weak.

Due to the general requirements to detect and identify concealed weapons carried by people at safe standoff distances using imaging equipment with reasonably small aperture sizes, it is assumed that imaging systems designed for this scenario would most likely operate in the submillimetre/THz regime, probably in one of the 320- or 640-GHz atmospheric transmission windows.

An imaging system shall meet resolution requirements to allow for enough spatial cycles on a subject at a desired standoff distance to enable an observer to adequately perform the task of either detecting or identifying a concealed weapon or other target against typical background materials. In addition, such a system shall be capable of capturing image frames with enough signal-to-noise ratios to produce good imagery at video frame rates. Figure A.2 illustrates how combining visual and submillimetre wavelength images helps to detect, locate and identify a concealed weapon under the clothing of the first and second person from the right.

Figure A.2 — Combined visual and submillimetre wavelength image of four subjects

A.2 Recommendations for full body video recording

Guidance concerning the positioning of the subject and camera is provided, as well as examples of lighting arrangements. The intention of this guidance is to ensure that the subject's full body is properly positioned and uniformly illuminated, thereby producing images that are conformant with this document and are without shadows or hot spots on any part of the body.

Distortion that is noticeable by human examination is forbidden. In particular, fish eye effects caused by wide angle lenses combined with too close camera placement shall not be present. Figure A.3 illustrates the relationship between camera FOV shown as 50 mm lens focal length, camera to subject distance and subject height.

A combined photo and video studio is typically a professionally operated facility, equipped with a digital camera, multiple adjustable light sources, a suitable background or backdrop cloth, and subject positioning apparatus designed to obtain high quality full body imaging.

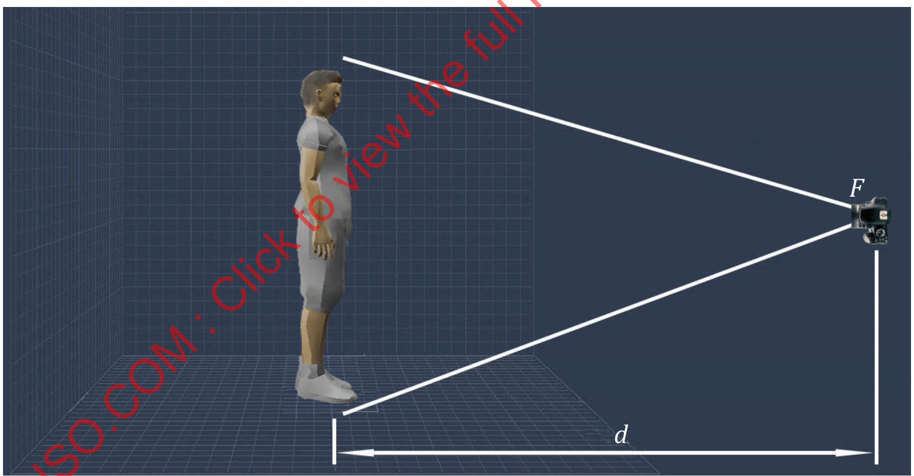
Vertical orientation of the camera is recommended for full body images and videos for three reasons:

— Vertical orientation allows the use of available pixels in an economical way.

— Vertical camera orientation is used for full body images.

— Video frame rotation is easily handled in the post-processing.

The use of a treadmill for small recording studio setup is recommended to achieve good quality gait recognition video recordings or gait image sequences.



Key

F 50 mm lens focal length

d 4 m distance

Figure A.3 — Positions of subject and camera

When planning the studio, take into consideration safety and security rules. The studio should be a place where everyone is and feels safe. Although it is impossible to completely eliminate the risk of violence and other events such as fire, it ought to be possible to reduce these risks to a minimum by implementing a proper set of procedures.

There is no certainty about which instances trigger sudden violent or aggressive behaviour in particular individuals. Architectural decisions are important from the safety point of view. For example, the subject shall always be located in a place that enables the operator to evacuate the studio room through a door behind themselves. This arrangement is particularly important in prisons, police stations and similar facilities.

A.3 Gait recording — technical recommendations

A.3.1 General

In practice, it is easier to arrange good working conditions for enrolment in a studio or office than in other environments. Poor enrolment and surveillance conditions result in the creation of low-quality biometric references and probes. Surveillance is easiest to arrange when it is possible to attend to the planning of the surveillance environment, for example, when a new airport is planned or an existing one is refurbished or expanded. Proper construction of the walkway is vital for gait recognition matching quality, although several methods are available to mitigate environmental error factors[22], [24], [25].

Among the factors affecting the gait, variation due to viewpoint change is considered as one of the greatest challenges. Several methods are in use to tackle the problems caused by cross-view and cross-walking conditions with different pre-processing approaches and convolutional neural network architectures. Limiting the cross-view and cross-walking conditions gives better recognition conditions regardless of the processing methods. Viewing angle differences under 20° between the reference and probe are recommended[23].

A.3.2 Error factor limitation

To get the best possible gait identification results it is necessary to limit known error factors. To reach this goal for vision-based gait recognition it is necessary to disentangle the identity-unrelated factors which alter gait appearances. Stabilization of these factors and optimization of the factors of the whole gait recognition system both affect the recognition quality of the whole system. For example, adapting to the walking speed is a stabilization method and limiting the walking speed to the optimum detection range of the system is an optimization method.

The known factors can be grouped into three separate factor groups, namely subject-related, device-related and environment-related.

Subject-related factors are:

— walking speed;

— clothing; and

— item carrying.

Device-related variable factors are:

— frame rates; and

— image resolutions.

Environment-related factors are:

— illumination conditions; and

— camera viewpoints.

Among these factors, the change of viewpoints is a meaningful factor. Cross-view variance is one of the key problems in video-related tasks. The performance of the gait recognition system drops drastically when the viewpoint changes if the system ignores cross-view variations.

A possible approach to gait-based human identification is via similarity learning by deep convolutional neural networks (CNNs). This can be achieved with labelled multi-view human walking videos to train deep networks to recognize the most discriminative changes of gait patterns which suggest the change of human identity[23]. Another approach for cross-view action recognition is the use of ‘virtual views’ that connect the action descriptors extracted from one or several probe view(s) to those extracted from a reference view[27]. A variation of this is the use of ‘virtual paths’ where each point on the virtual path is a virtual view which is obtained by a linear transformation of the action descriptor[28].

A.3.3 Walkway construction

Whether for enrolment or surveillance, environmental factors are critical to the quality of a gait video recording site. The following design rules should be followed:

— Straight walking path on the walkway for the subjects.

— No bumps and imperfections on the floor.

— No distractions on the walkway like message boards, signs or personnel.

— No textures, posters or additional objects on the floor or on the walls.

— Neutral grey matte materials on the floor and walls.

— Even lighting on the subject path. Outside lighting blocked with a roof if over 200 lux (a lux is equal to one lumen per m^2).

— Uniform lighting level should be approximately 500 lux or more, depending on the entry and exit environmental lighting.

— Controlled colour temperature so that the camera's white balance algorithm is able to adapt to the light.

The recording shall be colour neutral and correctly exposed. It is important that the boundary between the whole body and the background should be clearly identifiable about the entire subject. There should be no shadows visible on the background behind the body image. The background should be plain and should contain no texture containing lines or curves that could cause computer pattern recognition algorithms to become confused. Therefore, the background should be a uniform colour.

In enrolment studios, either the camera follows the subject or a treadmill is in use with a stationary camera. Treadmill speed should be set for normal walking speed of 5 km/h or lower for people with walking difficulties. Treadmills can pose a real balance challenge for an unaccustomed user. For this reason, an assistant is sometimes required to ensure the subject's safety.

The normal self-selected gait speed of a healthy subject aged 10-79 years is 1,3-1,5 m/s (4,7-5,4 km/h). People over 80 years may have slower natural gait speed in the order of 1 m/s (less than 4 km/h). Gait velocity for long walkways can be higher than for short walkways[19].

Follow the studio setup guidelines set in ISO/IEC 39794-16:2021, Annex C.

A.3.4 Camera arrangements

A.3.4.1 General camera arrangements

Enrolment and surveillance differ from each other in many respects. One important one is the camera view angle. In reference recording during enrolment, the basic view angle is lateral, i.e. side view. It is possible to achieve surveillance camera installation in a controlled environment using the same view angles as in enrolment. Figure A.4 shows the approximation of a submillimetre imaging detail resolution for a super-resolved image.

Figure A.4 — Front and side camera views

A.3.4.2 Enrolment camera arrangements

A.3.4.2.1 General

The arrangement of the cameras along the walkway should be able to capture the movement of the subject mainly in the sagittal (lateral or side view) plane, as in the single camera setup, and optionally in the frontal (front view) plane, as in the dual camera setup. A sagittal plane is an anatomical plane which divides the body into right and left parts. A frontal plane is parallel to the long axis of the body and perpendicular to the sagittal plane. A frontal plane separates the body into front and back (dorsal) portions.

A.3.4.2.2 Single-view enrolment

The single camera setup in enrolment resembles that used in taking the frontal facial images. The subject is walking from right to left in front of the camera. If a treadmill is in use, then the subject is walking in front of the camera, which is recording the side view of the walking subject.

A.3.4.2.3 Dual-view enrolment

The dual camera setup utilizes two cameras recording at the same time when the subject is walking. The subject is walking from right to left in front of the first camera and towards the second camera. If a treadmill is in use, then the subject is walking in front of the first camera, which is recording the side view of the walking subject, while the second camera is recording in front of the subject.

When only one camera is in use the subject is walking first from right to left in front of the camera. The second recording is taken when the subject is walking towards the camera. The order of the recordings is not significant.

A.3.4.2.4 Multi-view enrolment

The multi-camera setup utilizes several cameras recording at the same time when the subject is walking. When only one camera is in use the subject walks multiple times along a path that is rotated in 45° angles in front of the camera, starting from the side view camera 1 in Figure A.5.

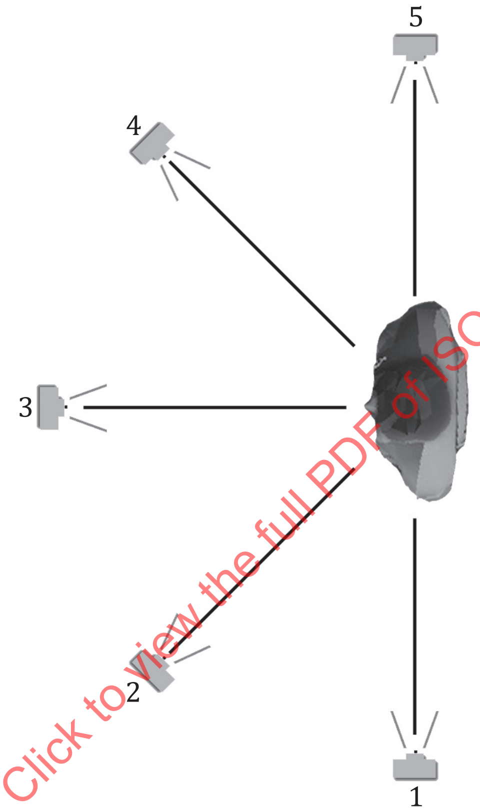


Figure A.5 — Multiple camera views

A.3.5 Studio lighting and layout

Lighting of the walkway or treadmill is achieved using several light sources. The main direction of the light is from the camera towards the subject. To minimize shadows on the background in long walkways it is advisable to use separate lights to illuminate the background.

Gait recognition walk-through video recordings can require more horizontal space than in Figure A.6, which gives minimum recommended measurements for the gait recording studio when using large surface light sources.

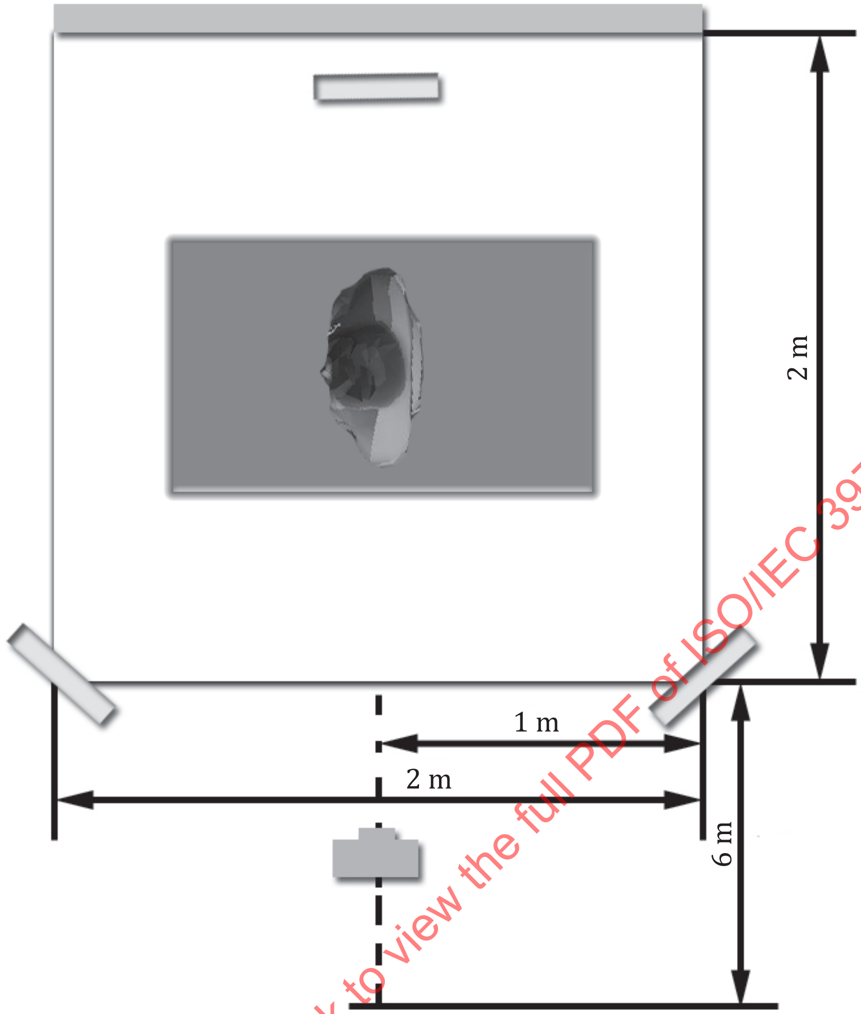


Figure A.6 — Minimum recommended measurements for the gait recording studio

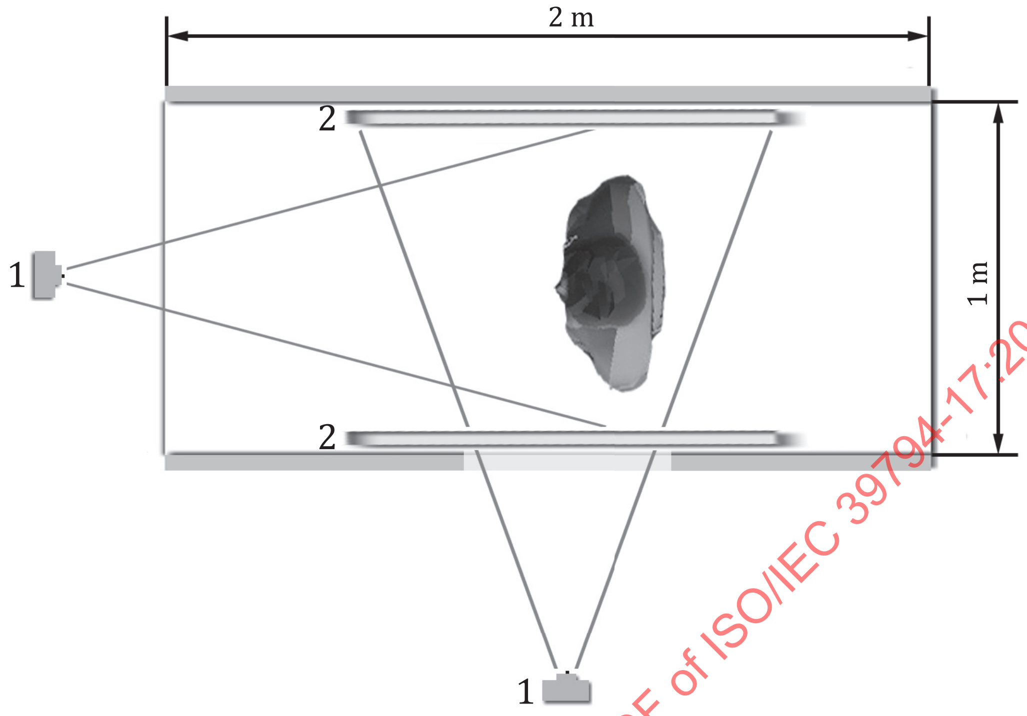
A.3.6 Walk-through gait recording corridor lighting and layout

The surveillance walk-through corridor for gait recording shall be constructed so that only one person at a time can walk by the camera. Background and foreground illumination may use strip lights, large illuminated light panels or indirect lighting to maintain even lighting and good contrast between the subject and background. Optimal corridor construct allows a full gait cycle to be recorded for all subjects walking through the surveillance corridor.

Corridor design shall be sufficiently flexible to meet changing needs and hardware. Subject safety and uninterrupted flow in the walk-through corridor should be considered with the location and position of any seating located near the corridor.

When the corridor is linked to an ABC gate then it is possible to automatically connect the subject's identification data with the imaging and recognition results of the corridor.

Figure A.7 illustrates the minimum recommended measurements and camera placements for the gait surveillance corridor when using large surface light sources.



Key

1 camera

2 light

Figure A.7 — Minimum corridor measurements and camera placements

Camera placements should be tested in advance using a mock up model. Short camera to subject distance (e.g. 2 m) and the use of uncorrected a wide-angle lens can lower the gait recognition efficiency.

Using both gait recognition and face recognition it is possible to build a walk-through access control system based on multimodal biometric verification. Facial video recording requires a second camera facing the object if sagittal gait recognition is in use[20].

A submillimetre recording device can require modifications to the corridor design shown in Figure A.7.

Annex B
(informative)

Encoding examples

B.1 Binary encoding examples

B.1.1 General

The gait image sequence data blocks used in these encoding examples are defined in the ASN.1 module in A.1. Data is encoded in DER. DER is a subset of BER providing for exactly one way to encode an ASN.1 value. DER is intended for situations when a unique encoding is required, such as in cryptography, and ensures that a data structure that needs to be digitally signed produces a unique serialized representation.

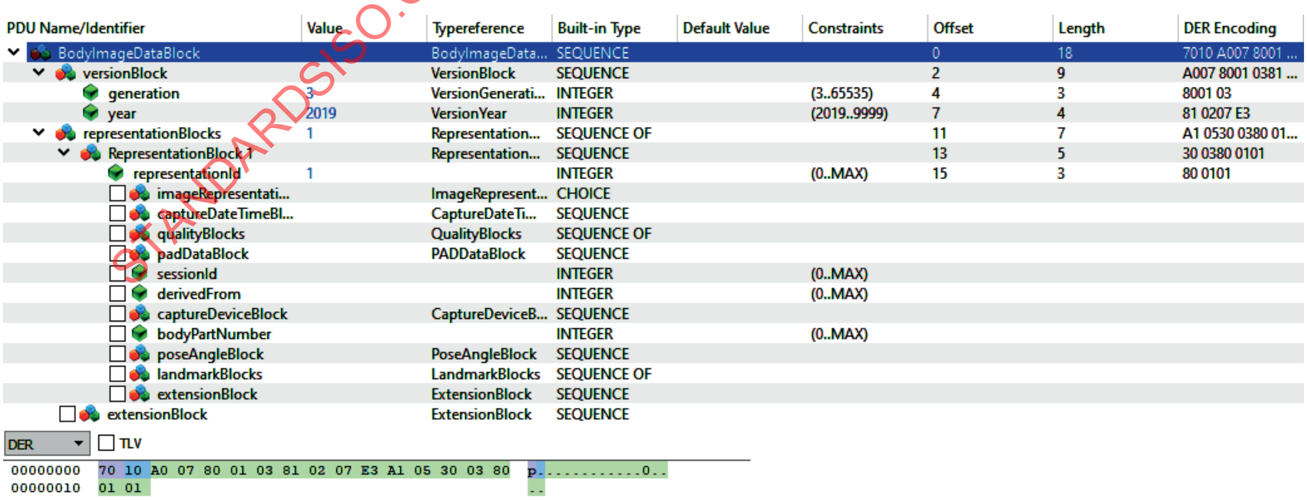
All ASN.1 schema elements are harmonized with the XML schema elements. Schema dependencies are described in more detail in B.2.

B.1.2 Binary encoding example using mandatory data fields

This data example structure for a gait still image sequence is compatible with the ISO/IEC 39794-16 data structure and data content. Its value is formally described below using ASN.1 notation.

```
value1 BodyImageDataBlock ::= {
  versionBlock {
    generation 3,
    year 2019
  },
  representationBlocks {
    {
      representationId 1
    }
  }
}
```

Figure B.1 shows the mandatory elements and the DER hexadecimal view at the bottom.



PDU Name/Identifier	Value	Typereference	Built-in Type	Default Value	Constraints	Offset	Length	DER Encoding
BodyImageDataBlock		BodyImageData...	SEQUENCE			0	18	7010 A007 8001 ...
versionBlock		VersionBlock	SEQUENCE			2	9	A007 8001 0381 ...
generation	3	VersionGenerati...	INTEGER		(3..65535)	4	3	8001 03
year	2019	VersionYear	INTEGER		(2019..9999)	7	4	81 0207 E3
representationBlocks	1	Representation...	SEQUENCE OF			11	7	A1 0530 0380 01...
RepresentationBlock 1		Representation...	SEQUENCE			13	5	30 0380 0101
representationId	1		INTEGER		(0..MAX)	15	3	80 0101
imageRepresentati...		ImageRepresent...	CHOICE					
captureDateTimeBl...		CaptureDateTi...	SEQUENCE					
qualityBlocks		QualityBlocks	SEQUENCE OF					
padDataBlock		PADDataBlock	SEQUENCE					
sessionId			INTEGER		(0..MAX)			
derivedFrom			INTEGER		(0..MAX)			
captureDeviceBlock		CaptureDeviceB...	SEQUENCE					
bodyPartNumber			INTEGER		(0..MAX)			
poseAngleBlock		PoseAngleBlock	SEQUENCE					
landmarkBlocks		LandmarkBlocks	SEQUENCE OF					
extensionBlock		ExtensionBlock	SEQUENCE					
extensionBlock		ExtensionBlock	SEQUENCE					

DER TLV

```
00000000  70 10 A0 07 80 01 03 81 02 07 E3 A1 05 30 03 80   p...........0..
00000010  01 01                                              ..
```

Figure B.1 — ASN.1 mandatory elements for a gait image sequence

An example encoding can be retrieved from https://standards.iso.org/iso-iec/39794/-16/ed-1/en/

B.1.3 Binary encoding example using JPEG image sequence elements

The second example contains several optional elements and an empty imageRepresentation2DBlock. Second example values are formally described below using ASN.1 value notation.

```
value1 BodyImageDataBlock ::= {
  versionBlock {
    generation 3,
    year 2019
  },
  representationBlocks {
    {
      representationId 1,
      imageRepresentationBlock representation : imageRepresentation2DBlock : {
        representationData2D '00'H,
        imageInformation2DBlock {
          imageDataFormat code : jpeg,
          imageKind2D code : gait2D
        }
      },
      captureDateTimeBlock {
        year 2020,
        month 7,
        day 24,
        hour 15,
        minute 47,
        second 56,
        millisecond 389
      }
    }
    {
      representationId 2,
      imageRepresentationBlock representation : imageRepresentation2DBlock : {
        representationData2D '00'H,
        imageInformation2DBlock {
          imageDataFormat code : jpeg,
          imageKind2D code : gait2D
        }
      },
      captureDateTimeBlock {
        year 2020,
        month 7,
        day 24,
        hour 15,
        minute 47,
        second 56,
        millisecond 790
      }
    }
  }
}
```

Figure B.2 illustrates this sample ASN.1 JPEG image sequence elements and the DER encoding data example for some of the most common entry fields showing typical placeholder test values as an example.

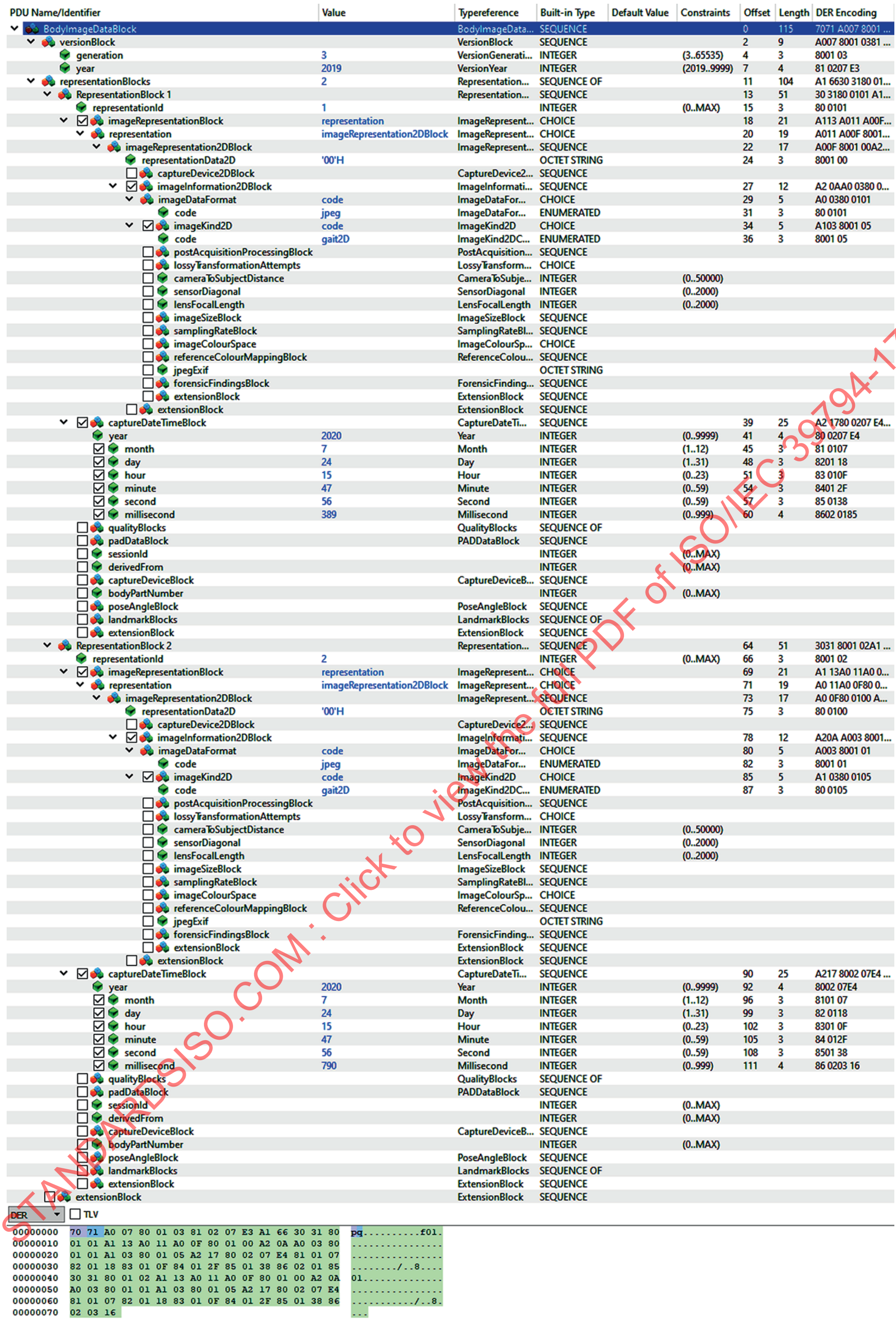


Figure B.2 — ASN.1 JPEG image sequence elements and the DER encoding data

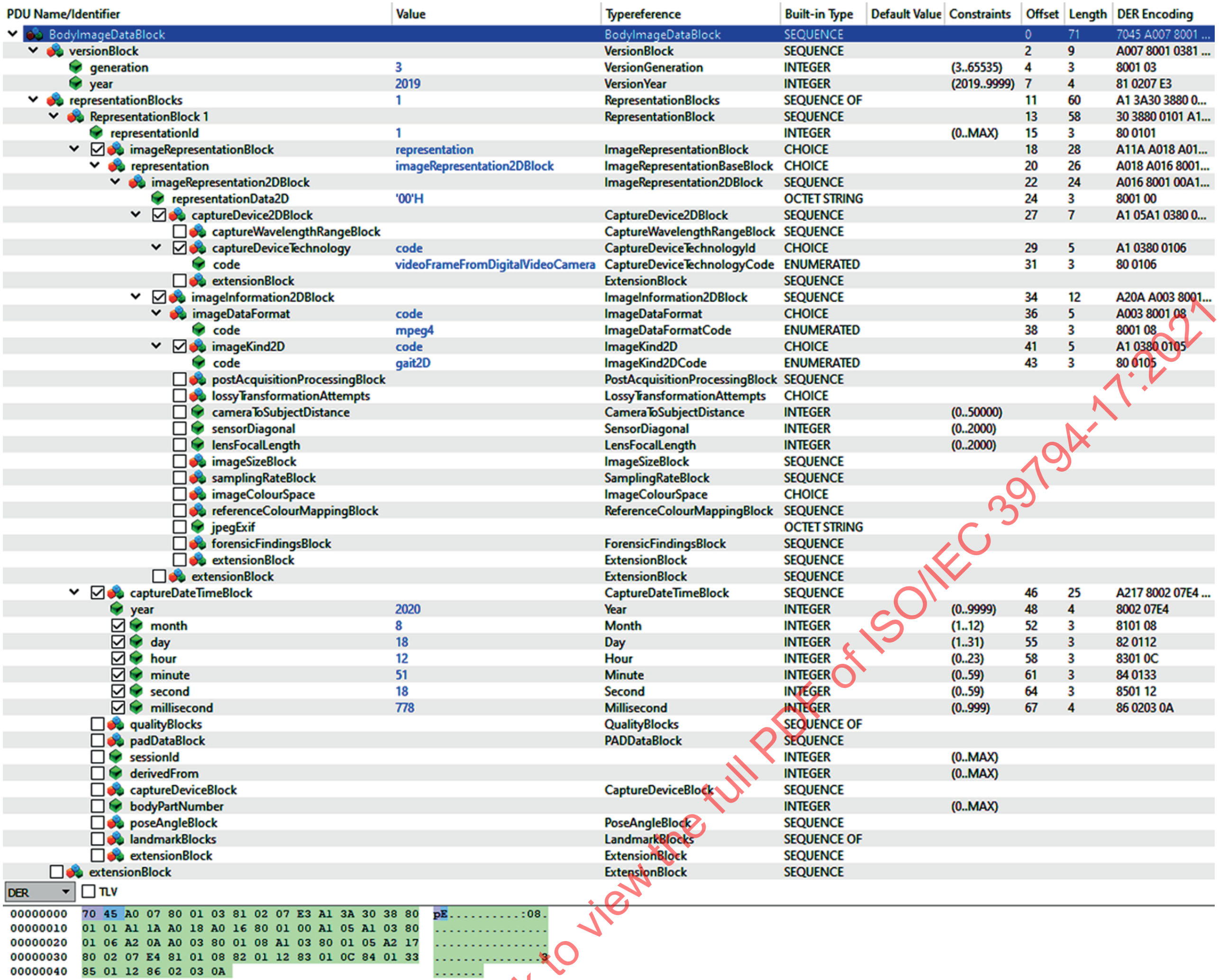
An example encoding can be retrieved from https://standards.iso.org/iso-iec/39794/-17/ed-1/en/

B.1.4 Binary encoding example using MPEG-4

MPEG-4 is one of the most common video formats available in digital cameras. The following example shows the minimum recommended data content for a gait MPEG 4 ASN.1 file.

```
value1 BodyImageDataBlock ::= {
  versionBlock {
    generation 3,
    year 2019
  },
  representationBlocks {
    {
      representationId 1,
      imageRepresentationBlock representation : imageRepresentation2DBlock : {
        representationData2D '00'H,
        captureDevice2DBlock {
          captureDeviceTechnology code : videoFrameFromDigitalVideoCamera
        },
        imageInformation2DBlock {
          imageDataFormat code : mpeg4,
          imageKind2D code : gait2D
        }
      },
      captureDateTimeBlock {
        year 2020,
        month 8,
        day 18,
        hour 12,
        minute 51,
        second 18,
        millisecond 778
      }
    }
  }
}
```

Figure B.3 illustrates this example ASN.1 MPEG-4 video image sequence elements and the DER encoding data example for some of the most common entry fields showing typical placeholder test values as an example.



PDU Name/Identifier	Value	Typereference	Built-in Type	Default Value	Constraints	Offset	Length	DER Encoding
BodyImageDataBlock		BodyImageDataBlock	SEQUENCE			0	71	7045 A007 8001 ...
versionBlock		VersionBlock	SEQUENCE			2	9	A007 8001 0381 ...
generation	3	VersionGeneration	INTEGER		(3..65535)	4	3	8001 03
year	2019	VersionYear	INTEGER		(2019..9999)	7	4	81 0207 E3
representationBlocks	1	RepresentationBlocks	SEQUENCE OF			11	60	A1 3A30 3880 0...
RepresentationBlock 1		RepresentationBlock	SEQUENCE			13	58	30 3880 0101 A1...
representationId	1		INTEGER		(0..MAX)	15	3	80 0101
imageRepresentationBlock	representation	ImageRepresentationBlock	CHOICE			18	28	A11A A018 A01...
representation	imageRepresentation2DBlock	ImageRepresentationBaseBlock	CHOICE			20	26	A018 A016 8001...
imageRepresentation2DBlock		ImageRepresentation2DBlock	SEQUENCE			22	24	A016 8001 00A1...
representationData2D	'00'H		OCTET STRING			24	3	8001 00
captureDevice2DBlock		CaptureDevice2DBlock	SEQUENCE			27	7	A1 05A1 0380 0...
captureWavelengthRangeBlock		CaptureWavelengthRangeBlock	SEQUENCE					
captureDeviceTechnology	code	CaptureDeviceTechnologyId	CHOICE			29	5	A1 0380 0106
code	videoFrameFromDigitalVideoCamera	CaptureDeviceTechnologyCode	ENUMERATED			31	3	80 0106
extensionBlock		ExtensionBlock	SEQUENCE					
imageInformation2DBlock		ImageInformation2DBlock	SEQUENCE			34	12	A20A A003 8001...
imageDataFormat	code	ImageDataFormat	CHOICE			36	5	A003 8001 08
code	mpeg4	ImageDataFormatCode	ENUMERATED			38	3	8001 08
imageKind2D	code	ImageKind2D	CHOICE			41	5	A1 0380 0105
code	gait2D	ImageKind2DCode	ENUMERATED			43	3	80 0105
postAcquisitionProcessingBlock		PostAcquisitionProcessingBlock	SEQUENCE					
lossyTransformationAttempts		LossyTransformationAttempts	CHOICE					
cameraToSubjectDistance		CameraToSubjectDistance	INTEGER		(0..50000)			
sensorDiagonal		SensorDiagonal	INTEGER		(0..2000)			
lensFocalLength		LensFocalLength	INTEGER		(0..2000)			
imageSizeBlock		ImageSizeBlock	SEQUENCE					
samplingRateBlock		SamplingRateBlock	SEQUENCE					
imageColourSpace		ImageColourSpace	CHOICE					
referenceColourMappingBlock		ReferenceColourMappingBlock	SEQUENCE					
jpegExif			OCTET STRING					
forensicFindingsBlock		ForensicFindingsBlock	SEQUENCE					
extensionBlock		ExtensionBlock	SEQUENCE					
extensionBlock		ExtensionBlock	SEQUENCE					
captureDateTimeBlock		CaptureDateTimeBlock	SEQUENCE			46	25	A217 8002 07E4 ...
year	2020	Year	INTEGER		(0..9999)	48	4	8002 07E4
month	8	Month	INTEGER		(1..12)	52	3	8101 08
day	18	Day	INTEGER		(1..31)	55	3	82 0112
hour	12	Hour	INTEGER		(0..23)	58	3	8301 0C
minute	51	Minute	INTEGER		(0..59)	61	3	84 0133
second	18	Second	INTEGER		(0..59)	64	3	8501 12
millisecond	778	Millisecond	INTEGER		(0..999)	67	4	86 0203 0A
qualityBlocks		QualityBlocks	SEQUENCE OF					
padDataBlock		PADDataBlock	SEQUENCE					
sessionId			INTEGER		(0..MAX)			
derivedFrom			INTEGER		(0..MAX)			
captureDeviceBlock		CaptureDeviceBlock	SEQUENCE					
bodyPartNumber			INTEGER		(0..MAX)			
poseAngleBlock		PoseAngleBlock	SEQUENCE					
landmarkBlocks		LandmarkBlocks	SEQUENCE OF					
extensionBlock		ExtensionBlock	SEQUENCE					
extensionBlock		ExtensionBlock	SEQUENCE					

DER TLV


```
00000000  70 45 A0 07 80 01 03 81 02 07 E3 A1 3A 30 38 80   pE..........:08.
00000010  01 01 A1 1A A0 18 A0 16 80 01 00 A1 05 A1 03 80   ................
00000020  01 06 A2 0A A0 03 80 01 08 A1 03 80 01 05 A2 17   ................
00000030  80 02 07 E4 81 01 08 82 01 12 83 01 0C 84 01 33   ...............3
00000040  85 01 12 86 02 03 0A                              .......
```


Figure B.3 — Example of the hierarchical display of theMPEG-4 ASN.1

An example encoding can be retrieved from https://standards.iso.org/iso-iec/39794/-17/ed-1/en/

B.2 XML encoding examples

B.2.1 General

The body image data blocks used in these encoding examples are of type BodyImageDataBlock defined in ISO/IEC 39794-16. All ASN.1 schema elements are harmonized with the XML schema elements. This document defines the profiles used for gait image sequence data. All elements are defined in ISO/IEC 39794-16. ASN.1 examples in ISO/IEC 39794-16 are fit for ISO/IEC 39794-17 (this document) application testing using single images.

Figure B.4 illustrates how gait profiles in this document and XSD dependencies are based on the ISO/IEC 39794-16 XSD element definitions and related to ISO/IEC 39794-5 and ISO/IEC 39794-1 XSD element definitions.

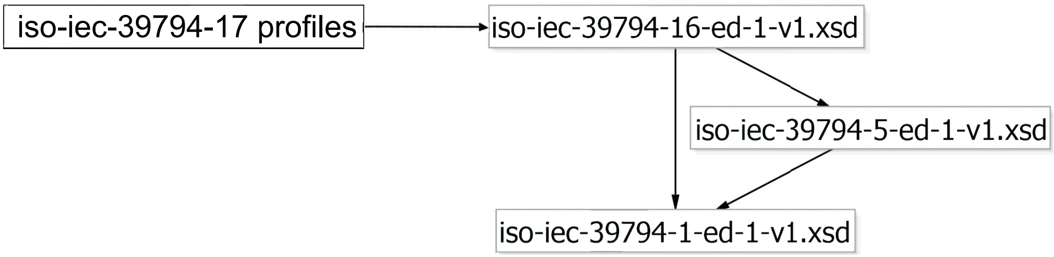


Figure B.4 — Gait profiles and XSD dependencies based on ISO/IEC 39794-16 XSD

Figure B.5 underlines how XSD modular structure is based on the main elements defined in the ISO/IEC 39794-16 XSD and available for the biometric modality and application profiles defined in this document.

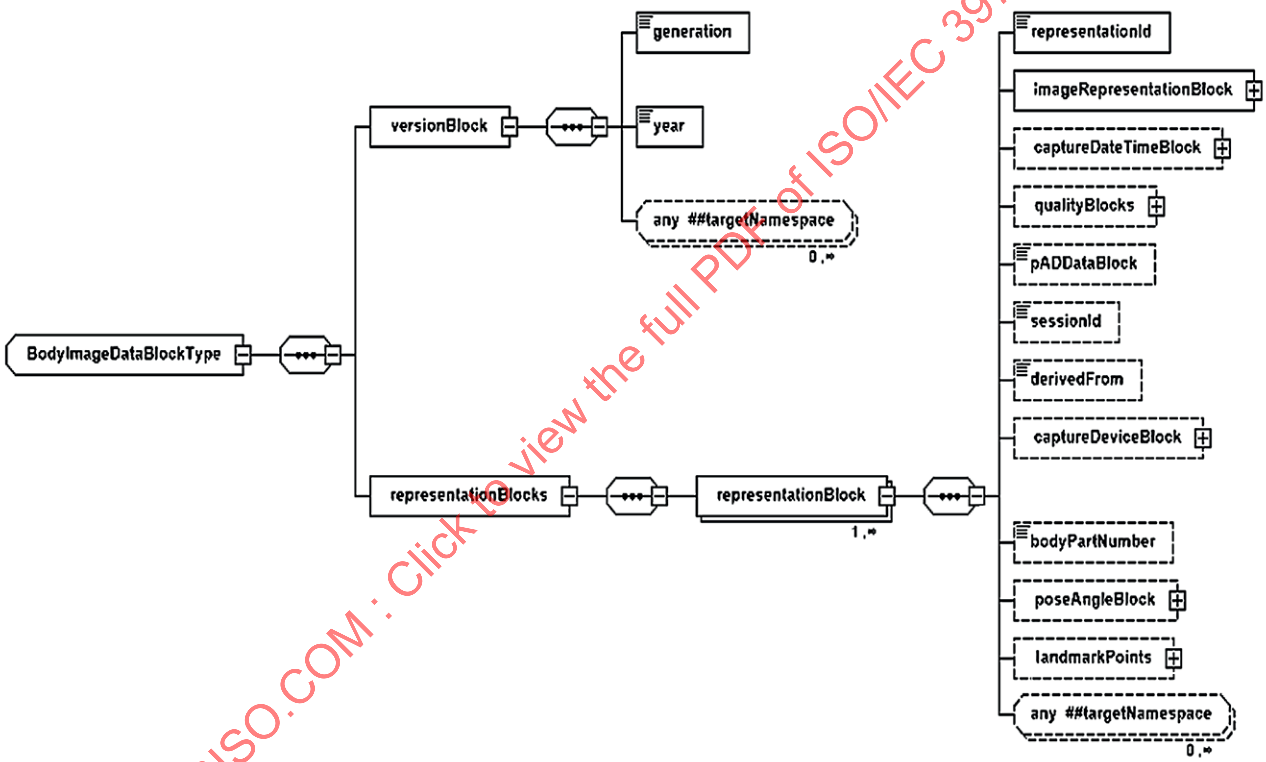


Figure B.5 — XSD modular structure is defined in ISO/IEC 39794-16 XSD

It is recommended to process any Motion JPEG gait image sequence into a set of separate JPEG images or to convert the sequence into a single MPEG-4 file. Motion JPEG (M-JPEG or MJPEG, .MJPG or .MJPEG file extension) is a video compression format in which each video frame or interlaced field of a digital video sequence is compressed separately as a JPEG image and stored in a single file.

The full examples containing the image data encoding can be retrieved from https://standards.iso.org/iso-iec/39794/-16/ed-1/en/ and for image sequence data encoding from https://standards.iso.org/iso-iec/39794/-17/ed-1/en/.

B.2.2 XML encoding example using mandatory data fields

This data example structure for a gait still image sequence is compatible with the ISO/IEC 39794-16 data structure and data content. Its value is formally described below using XML notation:

```
<?xml version="1.0" encoding="utf-8"?>
<!-- 39794-16 mandatory sample -->
<bim:bodyImageData      xmlns:cmn="https://standards.iso.org/iso-iec/39794/-1"
xmlns:fac="https://standards.iso.org/iso-iec/39794/-5" xmlns:bim="https://standards.iso.
org/iso-iec/39794/-16">
    <bim:versionBlock>
      <cmn:generation>3</cmn:generation>
      <cmn:year>2019</cmn:year>
    </bim:versionBlock>
    <bim:representationBlocks>
      <bim:representationBlock>
        <bim:representationId>1</bim:representationId>
      </bim:representationBlock>
    </bim:representationBlocks>
</bim:bodyImageData>
```


This data example structure for a gait still image sequence is compatible with the ISO/IEC 39794-16 data structure and data content. In Figure B.6, the mandatory XML elements are highlighted.

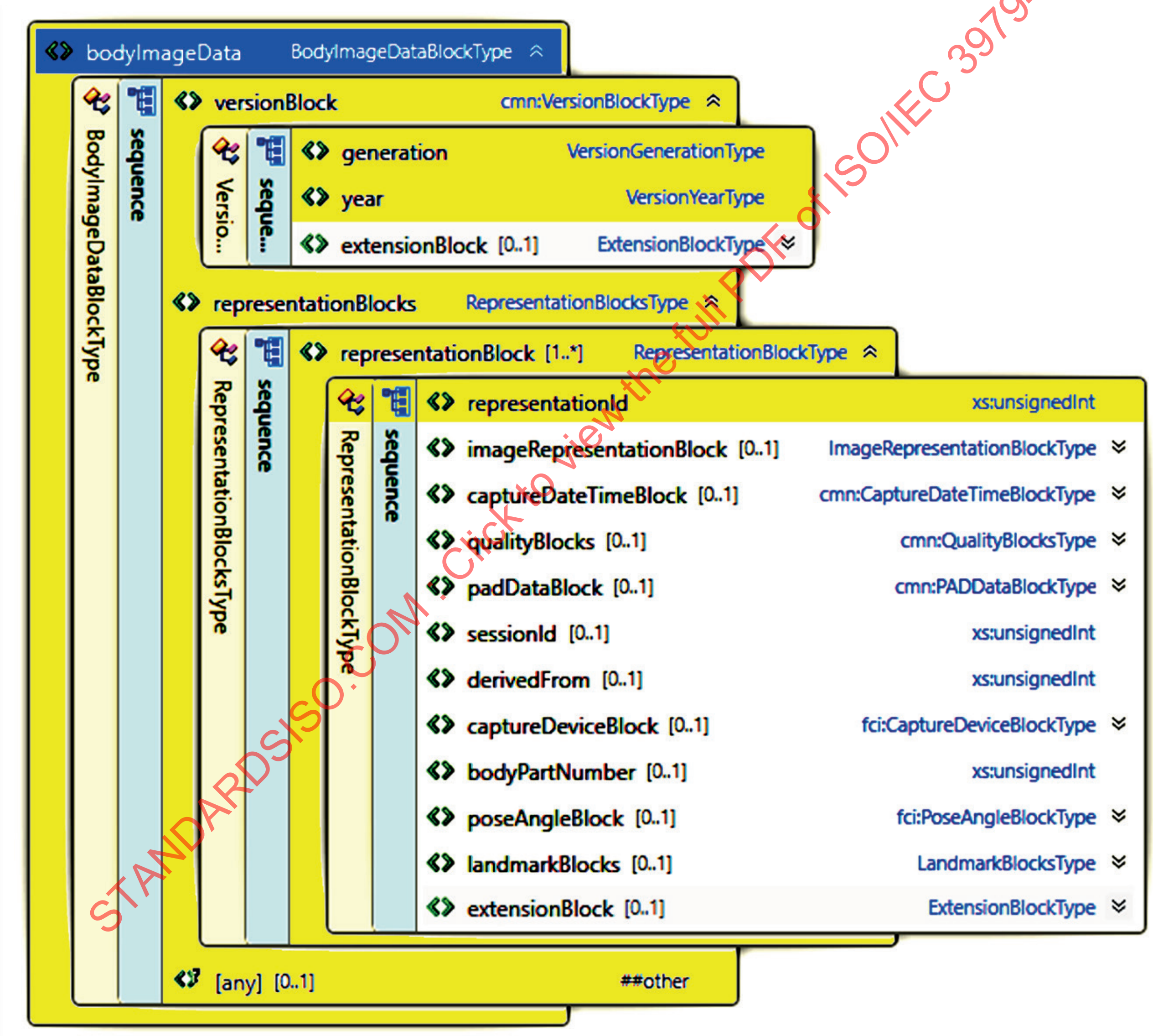


Figure B.6 — XML mandatory sequence structure

B.2.3 XML encoding example using JPEG gait sequence elements

This data example structure for a still image sequence is compatible with the ISO/IEC 39794-16 data structure and data content. Its value is formally described below using XML notation.:


```
<?xml version="1.0" encoding="utf-8"?>
<!-- 39794-17 JPEG sample -->
<bim:bodyImageData     xmlns:cmn="https://standards.iso.org/iso-iec/39794/-1"
xmlns:fac="https://standards.iso.org/iso-iec/39794/-5" xmlns:bim="https://standards.iso.
org/iso-iec/39794/-16">
  <bim:versionBlock>
    <cmn:generation>3</cmn:generation>
    <cmn:year>2019</cmn:year>
  </bim:versionBlock>
  <bim:representationBlocks>
    <bim:representationBlock>
      <bim:representationId>1</bim:representationId>
      <bim:imageRepresentationBlock>
       <bim:representation>
        <bim:imageRepresentation2DBlock>
          <!-- Gait image sequence image #1 -->
          <bim:representationData2D>AA==</bim:representationData2D>
          <bim:imageInformation2DBlock>
            <bim:imageDataFormat>
              <bim:code>
                <bim:jpeg>1</bim:jpeg>
              </bim:code>
            </bim:imageDataFormat>
            <bim:imageKind2D>
              <bim:code>
                <bim:gait2D>5</bim:gait2D>
              </bim:code>
            </bim:imageKind2D>
          </bim:imageInformation2DBlock>
        </bim:imageRepresentation2DBlock>
       </bim:representation>
      </bim:imageRepresentationBlock>
      <bim:captureDateTimeBlock>
        <cmn:year>2020</cmn:year>
        <cmn:month>7</cmn:month>
        <cmn:day>24</cmn:day>
        <cmn:hour>15</cmn:hour>
        <cmn:minute>47</cmn:minute>
        <cmn:second>56</cmn:second>
        <cmn:millisecond>389</cmn:millisecond>
      </bim:captureDateTimeBlock>
    </bim:representationBlock>
    <bim:representationBlock>
      <bim:representationId>2</bim:representationId>
      <bim:imageRepresentationBlock>
        <bim:representation>
         <bim:imageRepresentation2DBlock>
          <!-- Gait image sequence image #2 -->
          <bim:representationData2D>AA==</bim:representationData2D>
          <bim:imageInformation2DBlock>
            <bim:imageDataFormat>
              <bim:code>
                <bim:jpeg>1</bim:jpeg>
              </bim:code>
            </bim:imageDataFormat>
            <bim:imageKind2D>
              <bim:code>
                <bim:gait2D>5</bim:gait2D>
              </bim:code>
            </bim:imageKind2D>
          </bim:imageInformation2DBlock>
         </bim:imageRepresentation2DBlock>
        </bim:representation>
      </bim:imageRepresentationBlock>
      <bim:captureDateTimeBlock>
```

```
        <cmn:year>2020</cmn:year>
        <cmn:month>7</cmn:month>
        <cmn:day>24</cmn:day>
        <cmn:hour>15</cmn:hour>
        <cmn:minute>47</cmn:minute>
        <cmn:second>56</cmn:second>
        <cmn:millisecond>790</cmn:millisecond>
      </bim:captureDateTimeBlock>
    </bim:representationBlock>
  </bim:representationBlocks>
</bim:bodyImageData>
```


NOTE If JPEG is not used as a default value for the application or image data format is not detected after decoding then Image Data Format is included in the 2D Image Information block.

The structure of the ISO/IEC 39794-16 XSD element hierarchical structure is perhaps easiest to see by looking at the most frequently used elements for a sequence of two still images. A minimum full body image sequence BDB contains the data shown in Figure B.7.

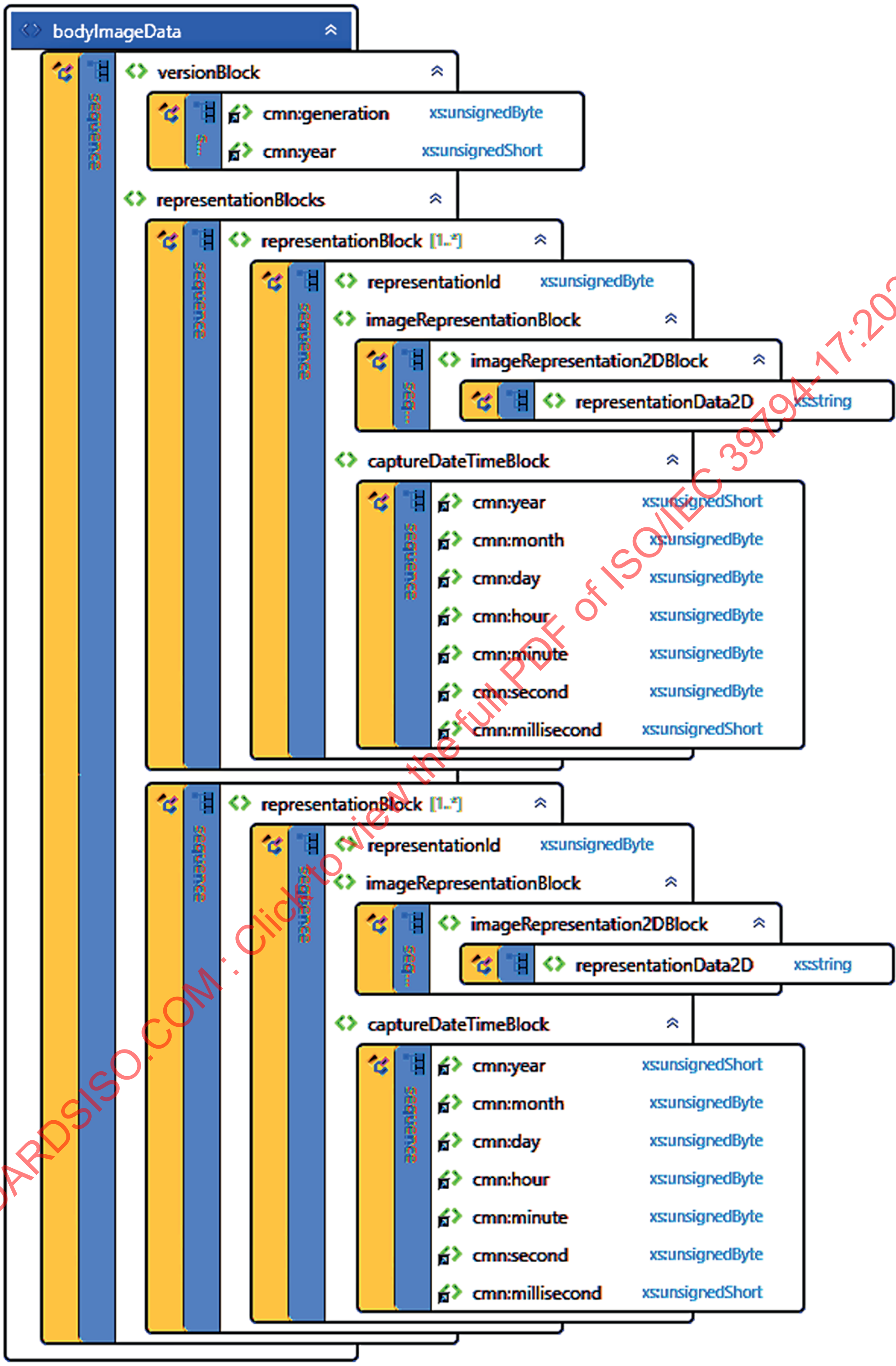


Figure B.7 — XML JPEG gait image sequence repeating structure

Note that the sessionId is not used if the representationId contains both the session identification and the representation identification information.

B.2.4 XML encoding example using MPEG-4

MPEG-4 is one of the most common video formats available in digital cameras. The following example shows the minimum recommended data content for a gait MPEG-4 XML file.

```
<?xml version="1.0" encoding="utf-8"?>
<!-- 39794-16 typical with image sample -->
<bim:bodyImageData      xmlns:cmn="https://standards.iso.org/iso-iec/39794/-1"
xmlns:fac="https://standards.iso.org/iso-iec/39794/-5" xmlns:bim="https://standards.iso.
org/iso-iec/39794/-16">
 <bim:versionBlock>
    <cmn:generation>3</cmn:generation>
    <cmn:year>2019</cmn:year>
  </bim:versionBlock>
  <bim:representationBlocks>
    <bim:representationBlock>
      <bim:representationId>1</bim:representationId>
      <bim:imageRepresentationBlock>
       <bim:representation>
        <bim:imageRepresentation2DBlock>
          <!--Gait mpeg4 video data in base64Binary sample value is 00H-->
          <bim:representationData2D>AA==</bim:representationData2D>
          <bim:captureDevice2DBlock>
            <bim:captureDeviceTechnology>
              <bim:code>

 <bim:videoFrameFromDigitalVideoCamera>6</bim:videoFrameFromDigitalVideoCamera>
              </bim:code>
            </bim:captureDeviceTechnology>
          </bim:captureDevice2DBlock>
          <bim:imageInformation2DBlock>
            <bim:imageDataFormat>
              <bim:code>
                <bim:mpeg4>8</bim:mpeg4>
              </bim:code>
            </bim:imageDataFormat>
            <bim:imageKind2D>
              <bim:code>
                <bim:gait2D>5</bim:gait2D>
              </bim:code>
            </bim:imageKind2D>
          </bim:imageInformation2DBlock>
        </bim:imageRepresentation2DBlock>
       </bim:representation>
      </bim:imageRepresentationBlock>
      <bim:captureDateTimeBlock>
        <cmn:year>2020</cmn:year>
        <cmn:month>8</cmn:month>
        <cmn:day>18</cmn:day>
        <cmn:hour>12</cmn:hour>
        <cmn:minute>51</cmn:minute>
        <cmn:second>18</cmn:second>
        <cmn:millisecond>778</cmn:millisecond>
      </bim:captureDateTimeBlock>
    </bim:representationBlock>
  </bim:representationBlocks>
</bim:bodyImageData>
```

In Figure B.8, MPEG-4 gait image sequence XML elements are highlighted.

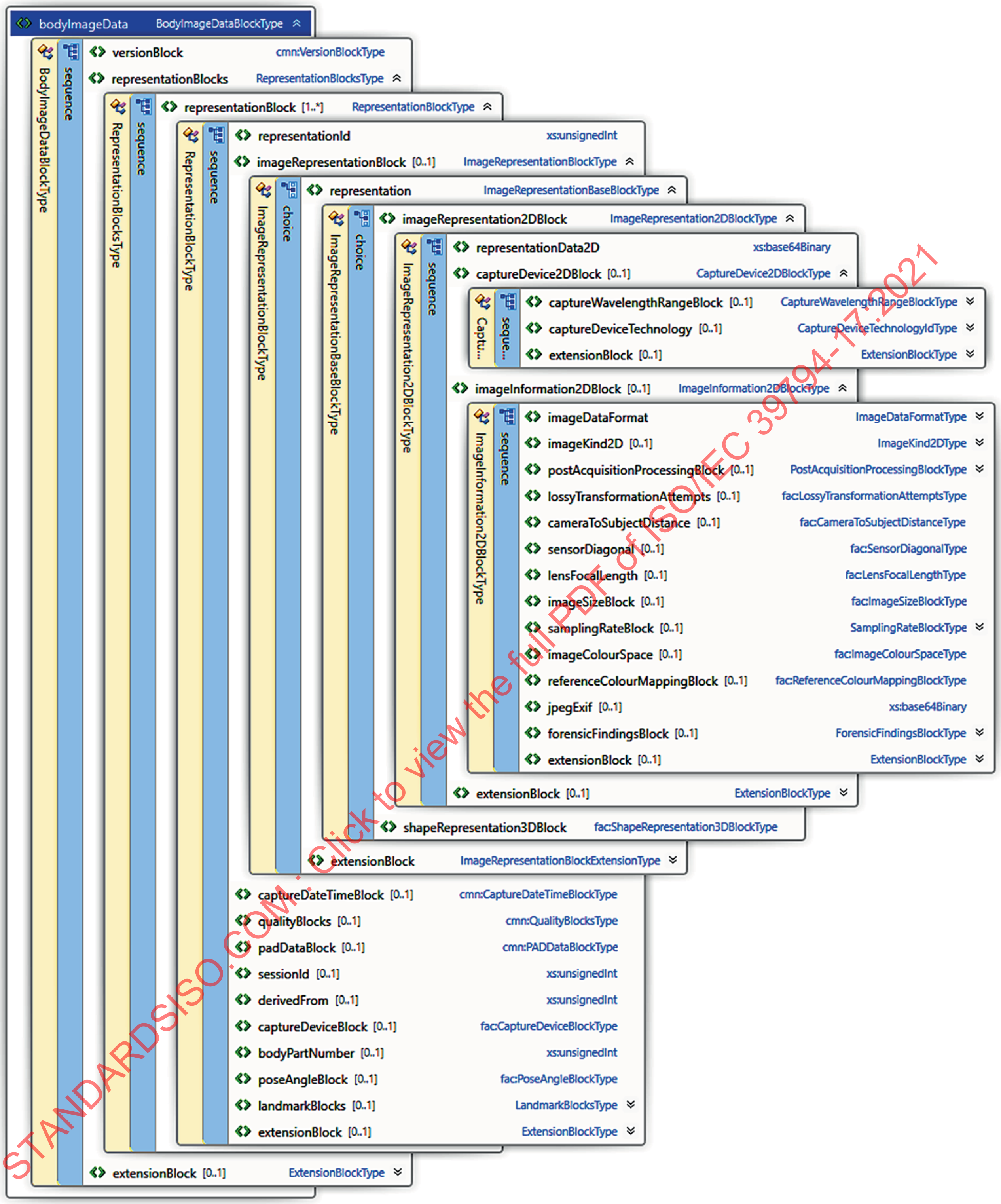


Figure B.8 — XML MPEG-4 gait image sequence structure

Annex C
(informative)

Image sequence acquisition measurements

C.1 General

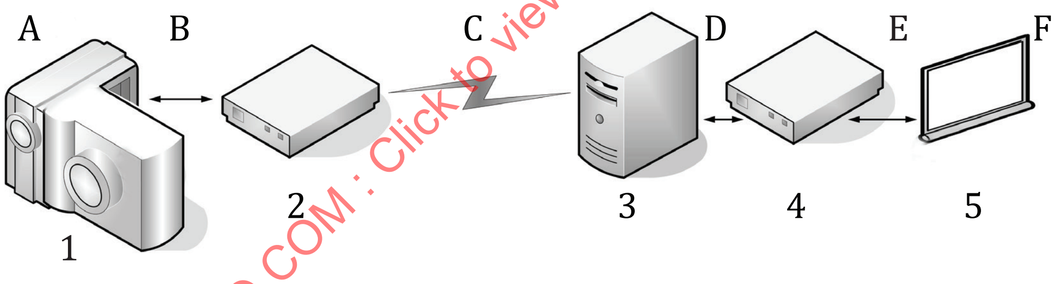
C.1.1 Introduction

The measurements described in the ISO/IEC 39794-16 are suitable for installation, adjustment, calibration and for maintenance purposes to maintain high image fidelity. Adjustment in biometrics is performed to set, for example, the camera and illumination so that images taken are conformant with the relevant specifications.

For advice on exposure metering and high-resolution camera image fidelity measurements, see ISO/IEC 39794-16.

C.1.2 Measurement environment

In video surveillance, the original video stream from camera sensor (A) is processed for output (B) and encoded for transmission (C) often inside the camera body. After the storage (D), a decoding step follows (E) before the video is displayed or processed by a recognition software (F). The H.264 and MPEG-2 standards define different video encoding (video compression) formats and transmission schemes. For this reason, the video transmission path can look different from the one shown in Figure C.1.



Key

A video
B camera output
C encoded data
D stored data
E decoded data
F data display or processing
1 access point
2 encoder
3 server
4 decoder
5 display

Figure C.1 — Video image measurement access points.

Video image measurements can be taken at several access points. Access point (1) shown in Figure C.1 should be used for measurements covered in this document.

In video surveillance applications, the probe images coming from the video cameras rarely meet all of the criteria set for the reference images. However, by controlling the image quality factors where possible, probe image quality is optimized. Image quality factors are affected by the video camera sensor and lens. These quality factors include resolution, noise (total, fixed pattern, and dynamic), dynamic range, exposure uniformity (vignetting) and colour quality. Lens distortion is the deformation of the image due to straight lines in the test target rendered as curved lines in the image from the camera. Lens distortion correction post-processing lowers the resolution of the image.

The measurements described in ISO/IEC 30137-1 are suitable for installation adjustment, calibration and for maintenance purposes to maintain high image fidelity in video surveillance systems. Adjustment in biometrics is performed to set the camera and illumination so that images taken are conformant with the relevant specifications, for example.

The gait image sequence enrolment studio camera performance measurements described in this document have been designed to be performed at moderate cost with moderately skilled operators. The tests do not require expensive or highly specialized equipment. They generally involve photographing standard targets under controlled lighting conditions and then analysing the resulting images on a computer.

Gait image sequences are processed to show either the subject's outlines or body landmark points. In practice, low resolution video frames contain enough information for the processing. However, good greyscale and colour fidelity help in the processing e.g. to remove background or find landmark points. If upper body movement recognition is in use, then it is advised to follow ISO/IEC 39794-16 measurement requirements.

C.2 Biometric capturing environment checking

C.2.1 General

The biometric capturing environment should be as stable as possible regarding the lighting and background. Basic checks for the camera should be performed before taking any video or a sequence of still images for the measurements described in this document. Camera lens and protective front window cleaning should be carried out as a preparation for the measurements. Clean the protective window on both sides. Clean the camera lens only if it is necessary, i.e. fingerprints on the lens or heavy dust on the lens surface. Lens cleaning paper tissues are a safe and inexpensive option for cleaning the camera lens.

C.2.2 Lighting checking

Subject and Scene Lighting: Lighting shall be equally distributed on the subject. There shall be no significant direction of the light from the point of view of the photographer.

The light falling onto (i.e. illuminating) and spreading over the subject is measured on several spots on the subject's body using a lux meter. The minimum illuminance value should be higher than 50 % of the maximum measured value. The measurements should be repeated on various spots on the walking path.

Hot spots and specular reflections: Hot spots (i.e. bright regions that result from light shining directly on subject) shall be absent. A single bare "point" light source like a camera mounted light is not acceptable for imaging. Instead, the illumination should be accomplished using other methods that meet the requirements specified in this subclause. Lighting artefacts can typically be avoided by increasing the angle between the lighting, subject and camera to 45°.

No unnatural colour: Colour photographs should use colour-balancing techniques such as continuous spectrum daylight illuminant. In demanding application imaging, fidelity[1] is determined by calculating and comparing the image CIE L*a*b*[2] Chroma tone values with the colour chart tone values as described in ISO/IEC 39794-16.

