INTERNATIONAL STANDARD

ISO/IEC 39794-17

First edition

Information technology — Extensible biometric data interchange formats —

Part 17: Gait image sequence data

PROOF/ÉPREUVE

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

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

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).

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.

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.

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 document 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 the 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.

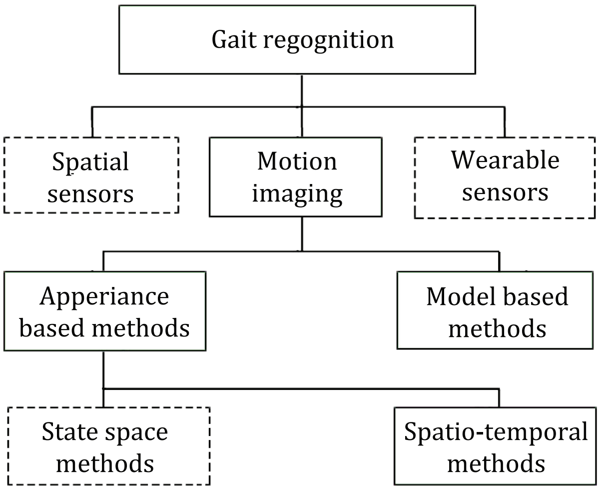


Figure 1 — Classification of gait recognition systems.

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.

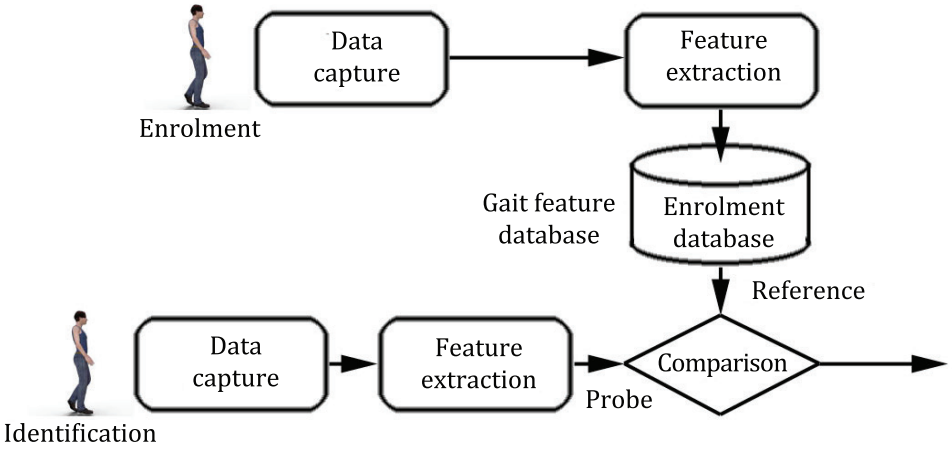


Figure 2 — Components of a gait image sequence biometric system.

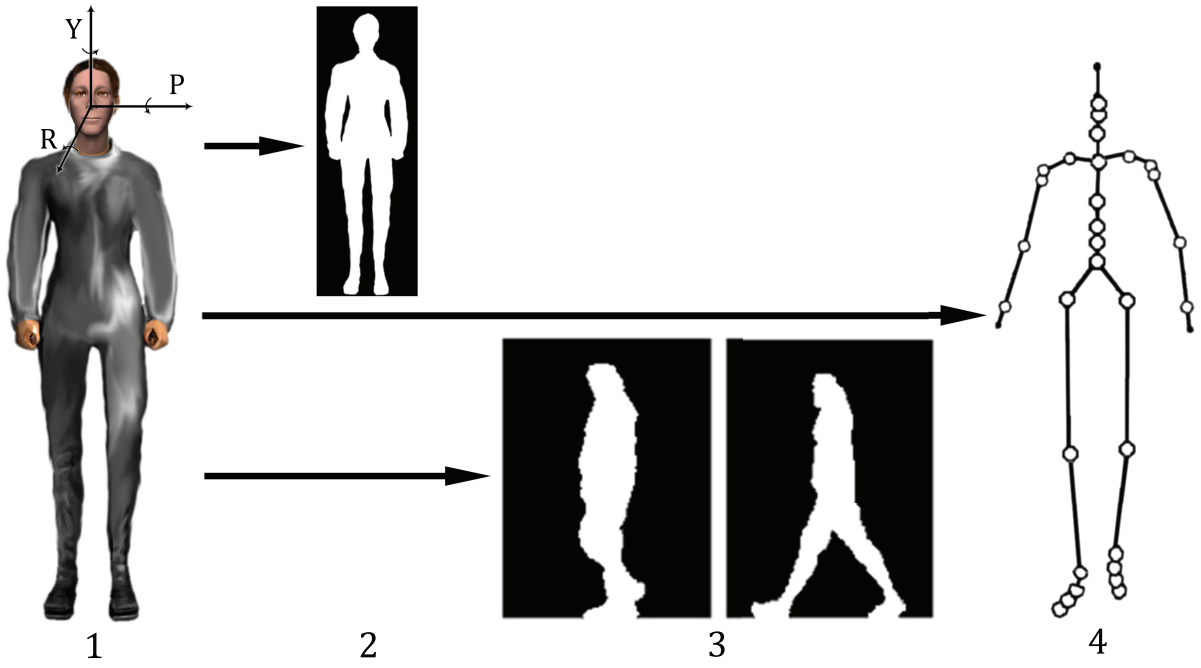
Comparison methods may use conventional feature-based template sets or deep convolution neural network (DCNN) feature vectors. After the feature vectors are generated using gait signatures and DCNN processing then the comparison is based on one of the many basic machine learning classification algorithms 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.

