INTERNATIONAL STANDARD

ISO/IEC 30137-4

First edition
2021-06

Information technology — Use of biometrics in video surveillance systems —

Part 4: Ground truth and video annotation procedure

Reference number
ISO/IEC 30137-4:2021(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 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 30137 series can be found on the ISO and IEC websites.

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.

Introduction

Considerable improvements in the performance of automated face recognition (AFR) have resulted in applications such as automated border controls, where facial images encoded in ePassports are compared with the face presented by a traveller at a control point. The success of these first generation AFR systems has encouraged suppliers to consider other applications, where the subject is not necessarily aware of the use of biometric comparison and where the environment for collection of images can be far from optimal. The inferior performance in such less-controlled identification applications can necessitate a greater involvement by trained personnel.

The ISO/IEC 30137 series provides guidance on the use of biometric technologies (primarily automated face recognition) in video surveillance systems (VSS) for several scenarios, including real-time operation against watchlists and post-event analysis of video data. The ISO/IEC 30137 series includes guidance on the selection and placement of cameras through to system specification, testing and maintenance. The ISO/IEC 30137 series uses the term VSS to replace the older but commonly used term, closed circuit television (CCTV).

The ISO/IEC 30137 series addresses the annotation of human beings. It is not intended to provide for annotation of non-human objects such as cars, animals, or luggage.

Records conformant to this document can be produced from video in either of the following ways:

— automatically, in which software analyses video and estimates quantities defined in this document, or

— manually, in which human reviewers annotate video with a goal of producing ground truth video annotation, which can be used by a receiving system (i.e. any service or device that decodes, interprets and uses standardized data).

This supports several applications, including:

— People counting:

  — stating of the number of people present in a location,

  — stating of the number of people traversing a given point or volume,

  — stating of population density (e.g. in crowds),

  — measurement of crowd densities,

  — performance of crowd behavioural analyses.

— Automated detection and tracking:

  — automated enrolment (addition) of subjects to a watchlist, exhaustively or after behavioural analysis,

  — detection of subjects, and parts of subjects (e.g. faces),

  — tracking of subjects through time, e.g. following motion in a single video,

  — tracking of subjects appearing through camera networks, including cases where a subject is viewed simultaneously by different cameras, and cases where the subject appears sequentially before several cameras,

  - re-identification, the process of connecting an identity of a subject across two or more video sequences.

- Automated identification:
  - in law enforcement, looking for subjects of interest present on watchlists (negative identification, blacklists),
  - in law enforcement, applications in review of post-event VSS video from one or multiple cameras against watchlists,
  - in private commercial settings, looking for individuals to be given preferential service,
  - identification of cooperative enrolled subjects (positive access control, whitelists).

This document includes annotation of the following information:

- Imaging type: single camera, sequential cameras, stereo cameras, combination, camera capture spectrum.
- When the subject appears in the video (start time) and when they leave (end time).
  - Brief description of the subject (what can be seen in the video?).
- Where and when the face of the subject appears.
  - Brief description of the face (pose, orientation, expression, occlusion).
- Intermediate tracking points between the start and end times, for subject and face.
- Absolute description of the subject:
  - estimated age, sex,
  - hair and eye colour,
  - estimated height and corpulence,
  - clothing and clothing colour,
  - glasses/hat,
  - best subject image or best subject face image.
- Subject interactions with other subjects and groups.
- Subject interactions with other video elements (bag, car, etc.).
- Known identity of the subject.
- The presence of other subjects who are not annotated.
- Regions of interest, outside of which an algorithm or receiving system would not operate.
- Absence: Where items of interest, including subjects, are known to be absent.

Standardized annotation supports evaluation, research and development, and operational deployment.

Information technology — Use of biometrics in video surveillance systems —

Part 4: Ground truth and video annotation procedure

1 Scope

This document establishes requirements for the annotation of humans, human faces and other body parts, and arbitrary objects appearing in imagery. It specifies the following:

— metadata to be inserted in a video stream;

— encoding of full and partial spatial and temporal ground truth information for:

  — objects present in a video, and

  — objects absent in a video;

— procedures for different annotation of known and unknown subjects.

This document does not specify:

— encoding of video data.

2 Normative references

There are no normative references in this document.

3 Terms and definitions

For the purposes of this document, the following terms and definitions 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 http://www.electropedia.org/

3.1
annotation
process of generating annotation data from imagery

3.2
annotation data
metadata associated with a subject traversing the field of view of a specific VSS camera

Note 1 to entry: An annotator preparing instances in accordance with this document should document the criteria under which a subject annotation was made. For example, it can be policy to not annotate faces for which interocular distance is below 12 pixels.

Note 2 to entry: If annotations are made by following a strict, tightly constrained or narrow set of criteria, then detection, tracking, recognition or algorithm is expected to be more accurate than if more permissive or general criteria has been used.

Note 3 to entry: An evaluation of, a tracking algorithm, for example, might exclude subjects that traverse in a non-conformant way. This could include factors such as the subject's direction of travel, obscuration by other people or objects, operational functionalities of the camera (such as correct focus) or environmental conditions (e.g. operation during night or day).

3.3
bounding box
rectangular region enclosing annotated object

Note 1 to entry: The major and minor axes of the rectangle are parallel to the edges of the images. For rotated boxes, the polygon annotation is to be used.

3.4
bounding polygon
arbitrary region enclosing annotated object

3.5
video surveillance system
system consisting of camera equipment, monitoring and associated equipment for transmission and controlling purposes, which can be necessary for the surveillance of a protected area

3.6
random access
ability to access arbitrary parts of a media item

3.7
recognition
process of assigning a biometric identifier to a subject

3.8
identification
process of determining a subject's identity by comparing imagery of a biometric mode against a database formed from imagery of individuals

Note 1 to entry: This generally does not include assigning an identifier when the target subject is not found in the database.

4 Abbreviated terms

AFR automated facial recognition

ROI region of interest

VSS video surveillance system

5 Conformance

A biometric data record conforms to this document if it satisfies all normative requirements related to:

— its semantic requirements,

— its encoding requirements for structure, data values, and the relationships between its data elements, as specified throughout Clauses 6 and 7 and Annex A for the biometric record format of this document, and

— the relationship between its data values and the input biometric data from which the biometric data record was generated.

6 Encoding of information supporting annotations

6.1 Overview

The following subclauses define encodings used in the full annotation of video clips, as detailed in Clause 7.

6.2 Region annotation

6.2.1 Content

An annotation of a body or body part shall enclose the region. An exception applies for a human face which may be annotated using anthropometric landmarks instead of, or in addition to, a bounding region.

A region annotation should be as precise as possible without adding an arbitrary margin around the object.

NOTE An object recognition algorithm can need to adjust the amount of spatial margin in the annotated region, depending on how it was trained and on its translational invariance.

If an object appears as two or more separated parts due to occlusion, two or more polygonal regions may be used. In this case, the polygonal regions shall be linked together by using a common identifier.

6.2.2 Encoding of a bounding box

Bounding boxes are the simplest mechanism for spatial annotation. They are rectangles whose major and minor axes are parallel to the image axes. They shall be encoded according to Table 1.

Table 1 — XSD schema for encoding of bounding box information


```
  <xs:import namespace="https://standards.iso.org/iso-iec/39794/-5"
schemaLocation="iso-iec-39794-5-ed-1-v1.xsd" />

  <xs:element name="regionData" type="RegionType">
    <xs:annotation>
      <xs:documentation>This is the root element of the 30137-4 data
structure.</xs:documentation>
    </xs:annotation>
  </xs:element>

  <xs:complexType name="RegionType">
    <xs:sequence>
      <xs:element name="bounding" type="BoundingType" />
      <xs:element name="objectClass" type="ObjectClassType" />
    </xs:sequence>
  </xs:complexType>

  <xs:complexType name="BoundingType">
    <xs:choice>
      <xs:element name="boundingBox" type="BoundingBoxType" />
      <xs:element name="boundingPolygon" type="BoundingPolygonType" />
    </xs:choice>
  </xs:complexType>

  <xs:complexType name="BoundingBoxType">
    <xs:sequence>
      <xs:element name="boundingBoxCoordinates" type="BoundingBoxCoordinatesType" />
      <xs:element name="localisationMethod" type="LocalisationMethodType" />
    </xs:sequence>
  </xs:complexType>
```

```
  <xs:complexType name="BoundingBoxCoordinatesType">
    <xs:sequence>
      <xs:element name="leftTopCoordinates" type="CartesianCoordinateType"/>
      <xs:element name="boxWidth" type="xs:unsignedInt"/>
      <xs:element name="boxHeight" type="xs:unsignedInt"/>
    </xs:sequence>
  </xs:complexType>

  <xs:complexType name="CartesianCoordinateType">
    <xs:sequence>
      <xs:element name="x" type="xs:int"/>
      <xs:element name="y" type="xs:int"/>
    </xs:sequence>
  </xs:complexType>

  <xs:complexType name="CartesianCoordinateListType">
    <xs:sequence>
      <xs:element name="cartesianCoordinate" type="CartesianCoordinateType"
maxOccurs="unbounded" />
    </xs:sequence>
  </xs:complexType>

  <xs:simpleType name="LocalisationMethodType">
    <xs:restriction base="xs:string">
      <xs:enumeration value="reviewer" />
      <xs:enumeration value="automatic" />
    </xs:restriction>
  </xs:simpleType>
```

6.2.3 Encoding of a polygonal region

Polygonal regions are the secondary mechanism for spatial annotation. They are available for annotation of objects that cannot be adequately localized, contained or demarcated by a bounding box. Bounding polygons shall be encoded according to Table 2.

Table 2 — XSD schema for encoding of polygon information


```
  <xs:complexType name="BoundingPolygonType">
    <xs:sequence>
      <xs:element name ="cartesianCoordinateList"
type="CartesianCoordinateListType" />
      <xs:element name="localisationMethod" type="LocalisationMethodType" />
    </xs:sequence>
  </xs:complexType>
```

6.3 Encoding of object class information

This subclause annotates arbitrary objects. This kind of object is referred to as its class. For biometric modalities, the class indicates the body part, such as a face, an ear, or a whole body. For other (generally) non-human objects, the class indicates a noun such as car or suitcase. The encoded data shall identify which body part or object is annotated according to Table 3. In cases where multiple modalities appear in one annotated region (e.g. face and ear), the encoded data shall represent at least one object. It supports annotation data of multiple objects.

Table 3 — XSD schema for encoding of object class information

```
  <xs:complexType name="ObjectClassType">
    <xs:sequence>
      <xs:element name="biometricModality" type="BiometricModalityType"
minOccurs="0" />
      <xs:element name="userDefinedObject" type="xs:string" minOccurs="0" />
    </xs:sequence>
  </xs:complexType>
```