International Standard

ISO/IEC 30137-1

Second edition
2024-03

Information technology — Use of biometrics in video surveillance systems —

Part 1: System design and specification

Technologies de l'information — Utilisation de la biométrie dans les systèmes de vidéosurveillance —

Partie 1: Conception et spécification

Reference number
ISO/IEC 30137-1:2024(en)

COPYRIGHT PROTECTED DOCUMENT

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).

ISO and IEC draw attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO and IEC take no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO and IEC had not received notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents and https://patents.iec.ch. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

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*.

This second edition cancels and replaces the first edition (ISO/IEC 30137-1:2019), of which it constitutes a minor revision. The changes are as follows:

— in the interest of using inclusive language, the terms "black list" and "white list" have been updated to "block list" and "allow list", respectively;

— minor editorial corrections have been made throughout the text, including corrections to cross-referencing within the document itself.

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 automatic facial recognition (AFR) technologies have resulted in applications such as automated border control using the facial images encoded in e-passports and implemented in systems whereby the identity of a co-operative traveller is verified in an environment designed for the collection of uniformly illuminated and optimally posed images. The success of these first generation AFR systems has encouraged suppliers to consider other applications where the environment for collection of images may be far from optimal. The inferior performance in such identification applications with less control can necessitate a greater involvement by trained personnel.

The ISO/IEC 30137 series provides guidance on the use of biometric technologies in video surveillance systems (VSSs), a framework for performance testing and reporting of such systems, and procedures for establishing ground truth and annotating video data for testing purposes.

This document provides the architecture, use cases and system design. The use cases include real-time alerting to the presence of individuals of interest, law enforcement applications such as reviewing post-event video footage from one or more cameras against pre-populated watchlists, commercial uses such as the identification of individuals who are to be given preferential service, and faces added to (enrolled in) a watchlist following observation of behaviours in the video material.

Other scenarios include measurement of crowd densities and determining numbers of individuals traversing a given point. While these are not the focus of this document, they are closely related and information on these scenarios is therefore included in Annex A.

Information technology — Use of biometrics in video surveillance systems —

Part 1: System design and specification

1 Scope

The ISO/IEC 30137 series is applicable to the use of biometrics in VSSs (also known as closed circuit television or CCTV systems) for a number of scenarios, including real-time operation against watchlists and in post-event analysis of video data. In most cases, the biometric mode of choice will be face recognition, but this document also provides guidance for other modalities, such as gait recognition.

This document:

— defines the key terms for use in the specification of biometric technologies in a VSS, including metrics for defining performance;

— provides guidance on the selection of camera types, placement of cameras, image specification, etc., for the operation of a biometric recognition capability in conjunction with a VSS;

— provides guidance on the composition of the gallery (or watchlist) against which facial images from the VSS are compared, including the selection of appropriate images of sufficient quality, and the size of the gallery in relation to performance requirements;

— makes recommendations on data formats for facial images and other relevant information (including metadata) obtained from video footage, used in watchlist images, or from observations made by human operators;

— establishes general principles for supporting the operator of the VSS, including user interfaces and processes to ensure efficient and effective operation, and highlights the need to have suitably trained personnel;

— highlights the need for robust governance processes to provide assurance that the implemented security, privacy and personal data protection measures specific to the use of biometric technologies with a VSS (e.g. internationally recognizable signage) are fit for purpose, and that societal considerations are reflected in the deployed system.

This document also provides information on related recognition and detection tasks in a VSS, such as:

— estimation of crowd densities;

— determination of patterns of movement of individuals;

— identification of individuals appearing in more than one camera;

— use of other biometric modalities such as gait or iris;

— use of specialized software to infer attributes of individuals, e.g. estimation of gender and age;

— interfaces to another related functionality, e.g. video analytics to measure queue lengths or to provide alerts for abandoned baggage.

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 terminology 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 Terms related to the target subject

3.1.1
operator
individual(s) responsible for day-to-day operation of the system

Note 1 to entry: This may include adjustment of the video surveillance cameras, selecting data suitable for use by the biometric application, and acting on the output of the biometric comparison process.

3.1.2
presentation attack
presentation of an artefact or of human characteristics to a biometric capture subsystem in a fashion that can interfere with the intended policy of the biometric system

3.1.3
target subject(s)
target(s)
individual(s) of interest

Note 1 to entry: A target subject is normally someone already enrolled in a *watchlist* (3.1.4). However, this is not always the case; in some scenarios they are a target because they are to be enrolled in a watchlist.

3.1.4
watchlist
list of individuals of interest (and their associated reference images) for detection by the video surveillance application

Note 1 to entry: The watchlist may be of individuals for whom an added service level is to be offered (e.g. VIPs or premium customers). This is sometimes referred to as an “allow list”.

Note 2 to entry: The watchlist may be a list of “wanted” individuals, e.g. individuals who should be denied access to premises or services. This is sometimes referred to as a “block list”.

Note 3 to entry: A system may have multiple watchlists of different groups of *target subjects* (3.1.3), and with different performance goals.

Note 4 to entry: In the case of target subject *back-tracking* (3.3.1), the watchlist normally contains only one target subject, or in the case of a group of individuals of interest, a few target subjects.

3.2 Terms related to VSS

3.2.1
codec
computer program capable of encoding or decoding a digital data stream or signal

3.2.2
compression ratio
measure of the compressed file size to that of the uncompressed file size

3.2.3
dropped frame
frame (3.2.4) from the video camera(s) that is not processed or is not available for facial detection and the creation of templates

Note 1 to entry: Normally measured in terms of either the number of frames per second dropped, or the percentage of the frames per second dropped.

3.2.4
frame
single image shown as part of a sequence of images in a video stream

3.2.5
frame rate
frequency (rate) at which an imaging device produces unique consecutive images called *frames* (3.2.4)

Note 1 to entry: Frame rate is normally expressed in frames per second (fps).

3.2.6
frame size
pixel dimensions of the *frame* (3.2.4) described in terms of horizontal and vertical pixels, and which can also be additionally described in terms of total megapixels

3.2.7
post-processing
steps performed after the biometric comparison process

EXAMPLE Triaging decisions based on a fusion of the quality and score metrics.

3.2.8
pre-processing
steps performed prior to the biometric comparison process

EXAMPLE Image quality enhancement, subject detection and feature extraction.

3.2.9
resolution
measure of the amount of detail that can be stored in an image

Note 1 to entry: Resolution is normally measured in pixels per millimetre.

3.2.10
subject tracking
process of aggregating multiple biometric samples for a single individual, possibly from multiple cameras, to avoid producing separate detection alerts for the same *target subject* (3.1.3)

3.2.11
video management system
VMS
component of a *video surveillance system* (3.2.12) that collects video from cameras and other sources, records that video to a storage device and provides an interface to both view the live video and to randomly access recorded video according to time

3.2.12
video surveillance system
VSS
system consisting of camera equipment, monitoring and associated equipment for transmission and controlling purposes, which may be necessary for the surveillance of a protected area

3.3 Terms related to biometric systems

3.3.1
back-tracking
act of finding the given image(s) of a face/individual by searching all video feeds where the individual can possibly have been seen

Note 1 to entry: It is possible, but not necessary, to use facial biometrics for back-tracking.

3.3.2
face detection
determination of the presence of faces within a video *frame* (3.2.4) and production of the location of each face in the frame

Note 1 to entry: Face detection is the first step in the face recognition process.

3.3.3
post-event analysis
non-*real-time analysis* (3.3.4) of data previously captured by video surveillance cameras

EXAMPLE To identify possible suspects following an incident or event.

3.3.4
real-time analysis
online processing of video surveillance data as it is captured

EXAMPLE To identify individuals held on a *watchlist* (3.1.4) so that immediate action can be taken.

3.3.5
Wiegand
de-facto wiring standard commonly used to connect a card swipe mechanism to the rest of an electronic entry system

3.3.6
zone of recognition
3-dimensional space within the field of view of the camera and in which the imaging conditions for robust biometric recognition are met

Note 1 to entry: In general, the zone of recognition is smaller than the field of view of the camera, e.g. not all faces in the field of view may be in focus and not every face in the field of view is imaged with the necessary inter-eye distance (IED).

3.4 Terms related to the environment and scenario

3.4.1
attractor
visual or acoustic cue within the environment which encourages individuals to look in a particular direction (i.e. towards the camera in a facial recognition application) in an attempt to improve recognition performance

3.4.2
choke point
point of congestion or obstruction through which individuals pass

3.4.3
lux
measure of illumination intensity

3.5 Symbols and abbreviated terms

AFIS	automated fingerprint identification system
AFR	automated facial recognition
APCER	attack presentation classification error rate
APNRR	attack presentation non-response rate
B&W	black-and-white
CCTV	closed circuit television (system); another term for video surveillance (system)
EXIF	exchangeable image file
FPS	frames per second
GUI	graphical user interface
HDR	high dynamic range
HMM	Hidden Markov models
IED	inter-eye distance; the distance (usually measured in pixels) between the centres of the eyes
IP	internet protocol
LFR	live facial recognition; real-time automated facial recognition using video surveillance cameras
MTF	modulation transfer function
NIST	National Institute of Standards and Technology
NPCER	normal presentation classification error rate
NPNRR	normal presentation non-response rate
OSDP	open supervised device protocol
PAD	presentation attack detection
PTZ	pan, tilt and zoom; a type of video surveillance camera that can be remotely adjusted (manually by the operator or automatically by using dedicated software)
SFR	spatial frequency response
SLI	standard lighting intensity
SNR	signal to noise ratio
SOP	standard operational procedure
VMS	video management system
VSS	video surveillance system

4 Comparison of terms used in biometric systems with those used in video surveillance

The video surveillance and biometrics communities both have well established vocabularies to describe the various components of a system, but the same term may sometimes be interpreted differently. While the

terms are defined in Clause 3, Table 1 highlights some of those terms and expressions where care needs to be taken when communicating with members of the video surveillance community.

Table 1 — Comparison of terms used in biometric systems with those used in video surveillance

Term	Definition within the context of automated biometric processing	Definition within the conventional use of human-led VSS, e.g. within the scope of IEC 62676 series
Crowd monitoring	Counting of individuals in a volume, or over a time interval	The observation of a group to determine collective behaviour or as part of a process to detect anomalous activity
Detection and localization	Biometric detection: the process of finding instances of a particular biometric mode, while correctly rejecting all instances of imagery not representing that biometric mode	Target detection: the process of finding targets of interest, such as humans or cars, in a video feed
Observation	Tracking: the process of spatially locating a particular biometric subject as it moves	Target observation: the process of following a particular target in a video feed
Recognition	The process for assigning a biometric identifier to a subject	The process of recognizing a familiar face; synonym for identification
Identification	The process of determining a subject's identity by comparing imagery of a biometric mode against a database formed from imagery of individuals. This generally includes not assigning an identity when the target subject is not present in the database	The process of a human determining a subject's identity using available (printed) galleries, or use of identity cues (clothing)
Verification	The process of confirming a subject's identity by comparing imagery of a biometric mode against a particular prior sample of a candidate individual	The process of confirming a target's identity
Inspection	Human review of the output from an automated biometric system to assess an alert from the biometric subsystem	Inspection: the detailed review of VSS imagery to determine more detailed information or characteristics, such as age or sex of an individual, brand of clothing, presence of jewellery
Alert	An indication that an identifier for an enrolled subject has been returned by the biometric recognition process	An indication issued by a camera, operator or system that an event of interest has occurred

5 Architecture

Figure 1 shows the process flow in a typical biometrically-enabled VSS with components such as the following.

1) Video surveillance cameras positioned to collect images in a form which supports comparison with images on the watchlist.

2) A VMS and infrastructure to organize and transmit footage from a number of cameras to the main server and storage system.

3) Software to detect and track faces (and/or other biometric features) in the video stream and to create biometric feature sets in the format developed by the supplier of the biometric recognition system. This can include feature sets created by combining features extracted from multiple face images from a single individual, continuously updated as new video frames are processed.

4) Comparison and decision software, again likely to be proprietary to the supplier of the biometric system, which determines whether the system has recognized an individual on the watchlist. The match criteria and decision thresholds may be different for groups of individuals on the watchlist, e.g. some can be considered low risk, with only minimal implications if they are not recognised by the system, whereas for others it can be imperative that they are recognized as soon as possible.

5) Alerts generated by the automated system are passed to the human operator for assessment.

6) An operator support environment to aid in making decisions on whether an alert should be followed up (and how) or rejected as a false alert.

7) Links to analytics systems to record the event and decisions taken, and to provide access to other information which can assist in disposal of the instance of recognition, e.g. previous instances of a similar match to the individual on the watchlist, and guidance on the appropriate action to be taken.

8) A systems management "bus" which enables configuration and operation of the key components in the biometric recognition system according to threat level, workload of human operators, time of day, etc., and which supports the merging of recognitions between cameras across the surveillance domain.

Figure 1 shows an example of a server-centric architecture. However, there are other models available, such as distributed architectures using edge computing (where part of the processing is done in the video camera of the VSS) or where cameras and computing resources are available within smart devices such as smartphones and PCs.

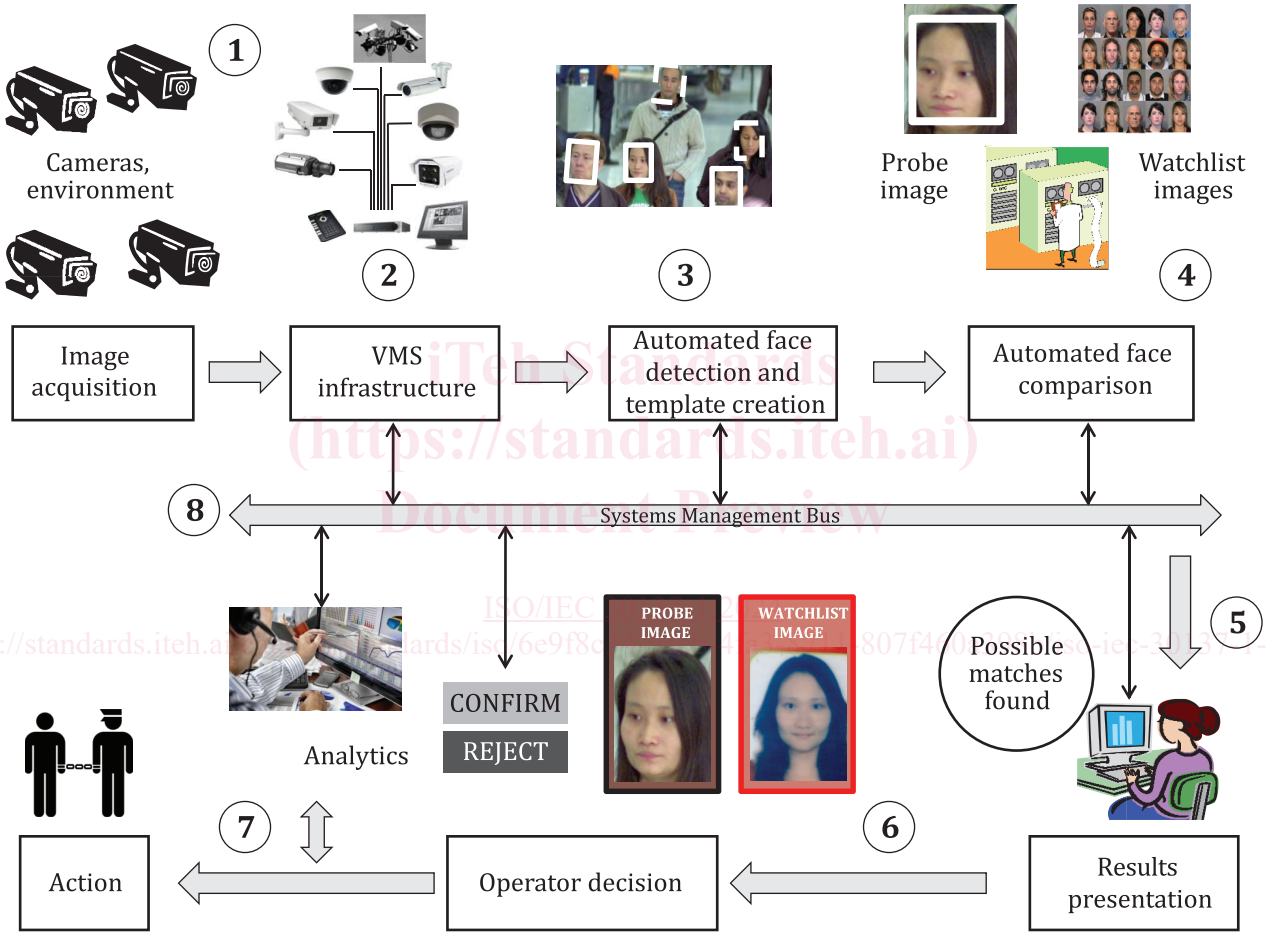

Figure 1 — Components of a biometrically-enabled VSS

6 Use cases

6.1 General

This clause provides examples of some of the different ways in which biometrics can be used in conjunction with VSS to support business needs across a range of organizations, including:

— police and law enforcement (and private security companies, such as those operating shopping malls and car parks) to alert to the presence of individuals of interest;

— police and law enforcement to manage the identification of individuals in video surveillance footage collected after a notable event or incident;

— commercial usage to alert to the presence of individuals of interest for whom special or differentiated levels of service are to be provided;

— commercial or government systems to manage the flow of individuals or queues, e.g. in accordance with agreed service levels;

— border services and client support organizations for quality assurance and customer support, e.g. following a complaint or an incident.

The use cases can be broken down into three broad categories, namely "post-event", "real-time" and "enrolment" applications (enrolment may be real-time or post-event). The following subclauses provide examples of some common use cases, described in terms of performance objectives and the roles played by various components of the system, including the responsibilities of the system operator.

6.2 Post-event use cases

In post-event use cases, the performance objective is the reliable detection, automated feature extraction, and searching of large numbers of target subjects against one or more watchlists or databases in an attempt to identify possible suspects, with a high probability that the candidate list returned by the biometric subsystem includes (at a high rank) those target subjects that have a matching template stored in the watchlist.

These use cases are challenging because in many cases the quality and positioning of the video cameras will be beyond the control of the operator of the biometric subsystem, and they will not have been installed with biometric applications in mind.

The operator normally has an “expert” role within the end-to-end process, selecting images suitable for submission as probes and examining candidates returned following a search of the database. They may be trained in facial comparison techniques, and the decision-making process may be supported by dedicated image analysis tools. In cases such as back-tracking or clustering (linking images of the same subjects together) the operator may also make use of other visual information (e.g. the individual’s clothes and relative location of cameras) to help them to confirm or refute potential matches.

Examples of post-event use cases include:

— post-event analysis of recorded video surveillance material (from one or more cameras) processed with the use of biometric recognition software to identify one or more individuals in frames or sequences (using one or more reference images);

— post-event analysis of recorded video surveillance material from more than one camera in which an individual (whether identified or not) is tracked (either forwards or backwards in time) and between cameras. This may involve more than just biometric applications, for example video analytics software;

— retrospective clustering — detecting and extracting faces from multiple sources of video for the purposes of clustering imagery sources of the same individual(s) together. This will normally need to be an automated process due to large numbers of subjects appearing in multiple video streams, although a human operator may subsequently review the results and intervene where they find subjects who have been wrongly classified.

6.3 Real-time use cases

In real-time applications, the performance objective is a high probability of the system alerting for target subjects with a matching template in a watchlist, and a low probability of an alert for subjects not in the watchlist. The watchlist will typically consist of a subset of images that are drawn from a larger image database and have been chosen to address a specific business objective.

These use cases are challenging because of the large amount of data that needs to be processed, especially if the system involves multiple cameras with multiple subjects in each frame. This presents a challenge