INTERNATIONAL STANDARD

ISO/IEC 30137-1

First edition
2019-05

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 30137-1:2019(E)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Fax: +41 22 749 09 47
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 http://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 Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 37, *Biometrics*.

A list of all parts in the ISO 30137 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.

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 less-controlled identification applications may necessitate a greater involvement by trained personnel.

The ISO 30137 series provides guidance on the use of biometric technologies in video surveillance systems (VSS), 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 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 30137 series is applicable to the use of biometrics in VSS (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 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;

— determining 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 other related functionality, e.g. video analytics to measure queue lengths or to alert 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 terminological databases for use in standardization at the following addresses:

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

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

3.1 Target subject related terms

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 could 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 will normally be 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* (3.1.3) (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 a "whitelist".

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 "blacklist".

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

Note 4 to entry: In the case of target subject *back-tracking* (3.3.1) the watchlist will normally contain only one *target subject* (3.1.3) (or in the case of a group of individuals of interest, a few target subjects).

3.2 VSS related terms

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 frames
frames from the video camera(s) that are not processed or are 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 described in terms of horizontal and vertical pixels, and which may 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 fusion of 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 Biometric system related terms

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

Note 1 to entry: Back-tracking may or may not use facial biometrics.

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-realtime analysis of data previously captured by video surveillance cameras

EXAMPLE To identify possible suspects following an incident or event.

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

EXAMPLE To identify individuals held on a watchlist 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 Environment/scenario related terms

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

CCTV Closed Circuit Television (system), another term for video surveillance (system)

FPS Frames per Second

LFR Live Facial Recognition, real time automated facial recognition using video surveillance cameras

GUI Graphical User Interface

HDR High Dynamic Range

IED Inter Eye Distance, the distance (usually measured in pixels) between the centres of the eyes

IP Internet Protocol

MTF Modulation Transfer Function

NIST National Institute of Standards and Technology

OSDP Open Supervised Device Protocol

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

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 listed above apply to this document, Table 1 below 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	For example, 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

Table 1 *(continued)*

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
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 identify 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

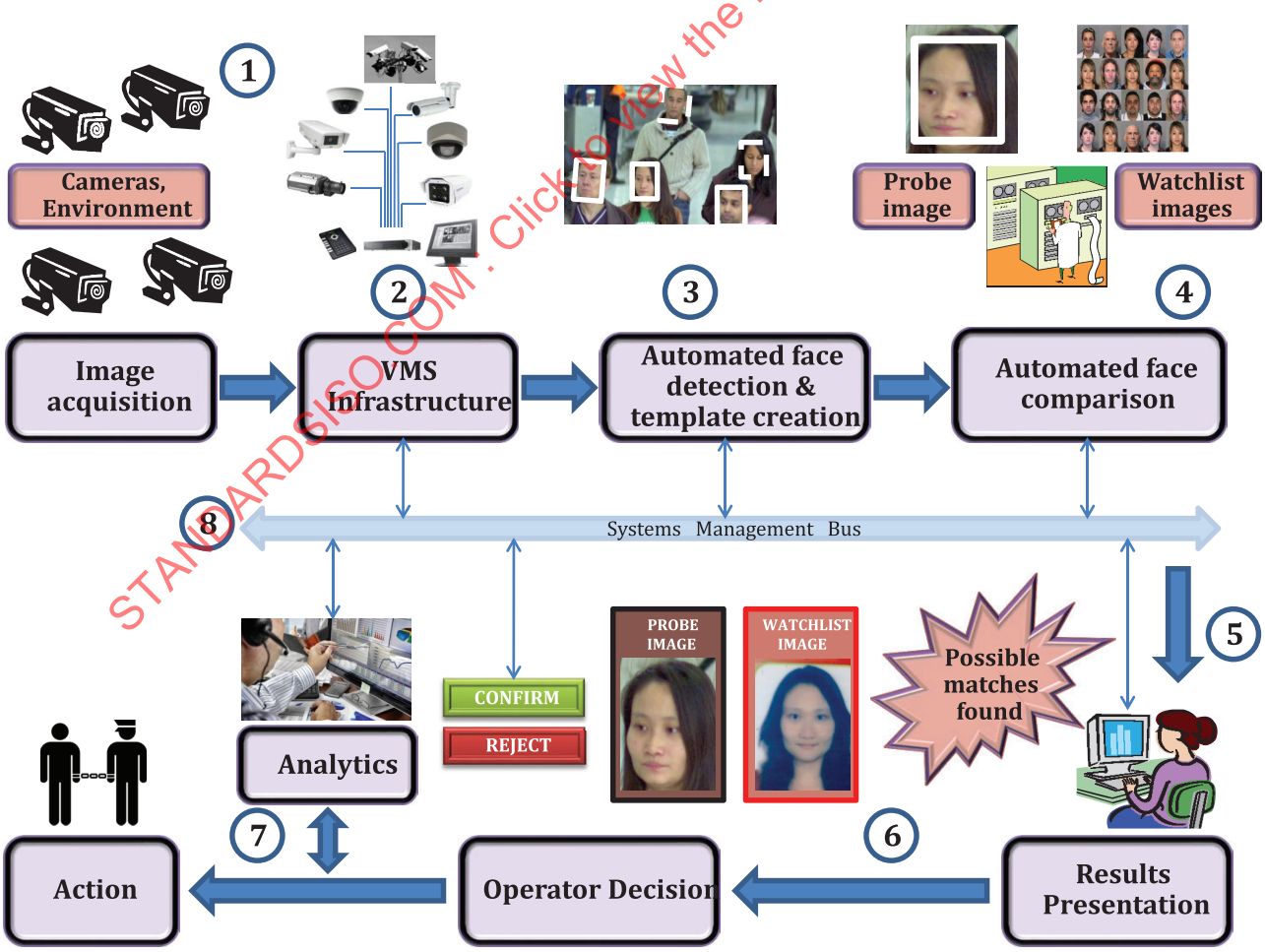


Figure 1 — Components of a biometrically enabled VSS

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

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 organise 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 recognised an individual on the watchlist. The match criteria and decision thresholds may be different for groups of individuals on the watchlist; e.g. some may be considered low risk, with only minimal implications should they not be recognised by the system, whereas for others it may be imperative that they are recognised 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 may 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 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.

6 Use cases

6.1 General

This section 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 organisations, 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 sections 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 backtracking 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 drawn from a larger image database, and that 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. Not only does this present a challenge in terms of search accuracy, it is vital that the end to end response time is fast enough to enable effective action to be taken when an alert is generated.

The role of the operator will typically be to assess any alerts from the biometric subsystem and to make an initial decision as to whether the alert is genuine or if it is a false match. They will usually also be responsible for instigating further action as appropriate, such as directing resources on the ground to detain or speak to the target subject in order to formally confirm their identity.

Examples of real time use cases include:

— alerting in real time (or near real time) to the presence of an individual traversing the field of view of a video surveillance camera, identified by the biometric subsystem as being someone whose

biometric data (e.g. a facial image) has previously been stored as a reference in a watchlist. For example, checking individuals entering a building or disembarking a plane or train against a watchlist, with the aim of either bestowing or denying particular privileges, and monitoring of video surveillance by law enforcement agencies for the purposes of crime prevention and public safety, This use case is sometimes referred to as Live Facial Recognition (LFR). A practical example illustrating the use of LFR can be found in Annex C;

— real time tracking of a particular individual of interest between the fields of view of a number of cameras, some of which may not overlap.

6.4 Enrolment use cases

In these use cases the goal is the successful enrolment of target subjects of interest into a database or watchlist, such that the biometric templates created are of sufficient quality for the intended use. Prior to enrolment, a biometric search may be carried out to determine if the target subject is already enrolled.

The operator may have a role in selecting the best quality images for enrolment, but in many cases the process will be fully automated. Machine learning and cognitive computing can be applied to help ensure that the best available images are selected for enrolment.

Examples of enrolment use cases include:

— enrolment (into a watchlist) of individuals who enter a protected zone or repeatedly visit the same area;

— "time clocking" individuals in situations where there is an interest in knowing how long they spend in a particular area, for example to monitor the time of service or queue length;

— enrolment of individuals traversing the field of view of a video surveillance camera into a database in order to support a watchlist application for future use within the same system, or to use in conjunction with other biometric applications and databases.

7 Specification of hardware and software

7.1 General

The IEC 62676 series already provides extensive information on camera selection, positioning, network bandwidth, performance considerations, storage requirements etc. for traditional (i.e. non-biometric) applications. This document therefore focuses on those aspects of the hardware and software components of the VSS that have a direct bearing on the performance of the biometric subsystem.

It is important to note that what may be an ideal set up for a conventional VSS may produce images that are very poorly suited for use in a biometric application. While the following recommendations are primarily applicable to an AFR system, they can also be adapted for other modes.

7.2 Physical environment

In many cases the environment in which the VSS is intended to operate will be beyond the control of those responsible for deploying or operating the cameras. Careful positioning of the cameras may help (see 7.5.2), but where it is also possible to exert some influence over the environment where the system operates, the following points should be considered:

— uneven floors and steps should generally be avoided as changes of angle/height often cause individuals to look down, thus making it hard to obtain usable images of the face;

— barriers may be introduced to modify the flow of individuals through the environment, ensuring they all pass through the field of view of the camera(s) at the correct distance and moving towards (in the case of an AFR system) the camera. Such techniques, together with careful positioning of the

cameras can increase the amount of time an individual is within the field of view which will in turn improve the performance of the system;

— choke points may be introduced to reduce the number of individuals passing through the field of view of the camera at any one time, thereby reducing the number of target subjects that need to be processed simultaneously by the biometric application. Choke points can also improve biometric sample collecting by limiting the speed with which target subjects move through a capture area, by improving the lighting at that location, and creating a situation where the pose of the target subject to the camera(s) is more favourable.

The introduction of barriers or choke points may have negative implications for individuals moving through the environment. Due consideration should be given to the need for usability, accessibility and user friendliness, and the balance between these factors and the need to obtain high quality images in order to maximise system performance should be determined on a case by case basis. See Annex B for more information on societal aspects to be considered when employing such techniques.

7.3 Illumination environment

Sufficient illumination is needed to support biometric processing. When possible, the following points should be considered:

— areas near windows or in sunlight should generally be avoided as the lighting cannot be controlled and will vary with the time of the day/year and prevailing weather conditions. Shaded or artificially illuminated areas generally produce better results;

— additional lighting may be introduced to raise overall light levels and also to ensure balanced illumination across the faces, with no strong shadows or excessively bright highlights; additional lighting also allows faster shutter speeds to be used, helping to avoid motion blur in the video;

— near-infrared lighting may also be used (in conjunction with surveillance cameras that can detect those wavelengths) to help reduce shadows and to improve biometric sample collection under low light conditions.

7.4 Inducing frontal view

AFR is highly sensitive to the orientation of the head relative to the optical axis of the camera. It is often possible to modify subject behaviour by installing “attractors” that encourage people to look in a particular direction, e.g. upwards or towards specific camera position(s).

When possible, the environment should also be inspected to determine if any sources of distraction exist, e.g. for example an extraneous television screen that could undermine the biometric capture process by adversely taking attention from the intended direction of view.

7.5 Cameras and supporting infrastructure

7.5.1 Selection of cameras

Camera and lens combinations should be selected such that the image resolution, frame rate, field of view and low-level light performance are capable of providing images of sufficient quality for use in the intended biometric application.

Several different quantifiable metrics are necessary to assess the performance of a video camera and associated system producing the video stream for video surveillance.

The spatial resolution of the video camera is one of the most important factors in determining the quality of an image captured by a VSS. A measure of spatial resolution is the MTF. The camera’s original image‘s MTF20 should be at least 0,4 cycles per pixel. The original image is the same as described in Annex D and refers to the unencoded signal.

In the case of IP cameras the measure of spatial resolution is impossible, because only an encoded signal is generated.

NOTE 1 ISO 12233 specifies methods for measuring the resolution of a video frame. It is applicable to the measurement of both monochrome and colour cameras which output digital data. Further information about the relationship between MTF and resolution can be found in Annex D.

NOTE 2 IEC 61966-8 is applicable to the characterization and assessment of a VSS for the purpose of colour management.

The exact requirements for the resolution of facial images captured by the camera varies with use case, the specific scenario and AFR algorithm being used, but current systems typically require a facial image that has an Inter Eye Distance (IED) of at least 50 pixels if they are to return useful results from a watchlist search.

Higher resolution images generally yield better results but the performance of many current algorithms plateaus at an IED of around 95 pixels, beyond which other factors such as camera position, lighting etc. may play a greater role than the image resolution in determining the performance.

NIST's "FIVE" study[12] provides more detailed information on the effect of image resolution on performance for a number of current AFR algorithms. The study also notes that many of the tested algorithms would not acquire faces with less than 20 pixels between the eyes, while others ceased to work if the IED was greater than 80 pixels.

As new algorithms are continually being developed, and in light of the interdependencies between image resolution and the various different algorithms, it is not possible to provide more specific guidance on "optimum" image resolution in this document.

Shutter speed, frame rate and image compression algorithms also have a bearing on image quality, the aim being to obtain the best possible image(s) from a target subject who may be moving across the field of view and who may only look briefly in the direction of the camera. In low light levels, individual frames may exhibit motion blur, reducing the amount of detail available in the image.

Consideration should be given to the focal length of the camera lens. This is dependent on the distance between the target subject and the camera but ideally should be similar to that used for the images stored in the watchlist, which therefore varies with the application. Large differences between the focal length used for the watchlist images and that of the video surveillance camera cause differences in perspective that may negatively impact on the performance of the biometric subsystem. In addition, optical distortion resulting from the type and quality of the lens may also have an adverse effect.

Many security cameras use a proprietary video format or container requiring specialized software from the manufacturer to play back the video stream. In some cases, the manufacturers do not provide the ability to transcode their proprietary format to more common formats such as H.264, thus requiring the development or procurement of additional software for this purpose. When selecting cameras consideration should be given to the need to transcode the recording format to a more common format.

The selection of appropriate video cameras plays a critical role in the overall performance of the system, but it is only one of many contributing factors. It is strongly recommended that organisations considering the introduction of a VSS with an AFR capability seek specialist advice to determine the most appropriate solution, and undertake proper performance testing both during the procurement stage and prior to operational deployment.

7.5.2 Positioning of cameras

In many existing video surveillance installations the cameras are positioned to monitor a wide field of view and as such are generally not well suited to biometric applications such as AFR. If video surveillance cameras are to provide images of a quality suitable for AFR systems, it is important to take particular care over their placement and orientation.

Five basic types of video surveillance environments of increasing complexity (and difficulty) are recognized[9]:

1. “Stationary” — as at passport control or biometric kiosk;
2. “Portal” — as in a one-way corridor or choke-point portal;
3. “Corridor” — as in a two-way corridor with more than one individual at time;
4. “Halls” — as in airport halls, shopping malls;
5. “Outdoors” — all other scenarios.

Example images from operational airport surveillance cameras corresponding to scenarios 1, 3 and 4 are shown in Figure 2 below. Note that the images depict the environments but not necessarily the optimum field of view of the camera.

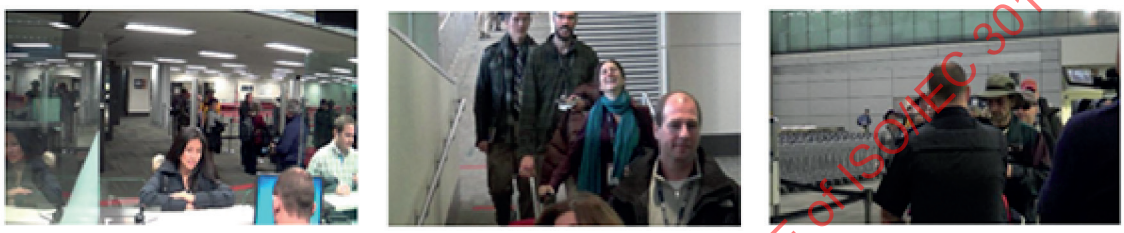


Figure 2 — Examples of Types 1,3 and 4 scenarios[9]

The following clauses provide guidance on positioning cameras in order to capture images suitable for use with AFR and for gait analysis.

7.5.2.1 Camera positioning for use with AFR

It is recommended, where possible, to use the Type 1 setup where the individual spends some time on the same spot, and on which the camera has been pre-focused. When the Type 1 setup is not possible, two solutions are recommended for the other scenarios:

1. using an array of video cameras;
2. using a combination of point-and-shoot cameras with VSS cameras running video analytics (which track a face, while measuring its resolution).

However, regardless of the setup type the objective should be:

— to obtain an optimised image (usually of the face) by minimising the pitch, yaw (rotation) and tilt angles, but without setting the camera so low down that faces of target subjects are obscured by individuals in front of them;

NOTE ISO/IEC 19794-5 “best practice” for facial image capture recommends a maximum of ±5°, although such stringent conditions are rarely met in video surveillance applications;

— to have the camera facing the direction of travel, and constrained to a choke point to avoid individuals moving across the main flow of traffic;

— to limit the number of individuals in the field of view at any one time (to avoid placing excessive load on the AFR software) whilst ensuring the area is adequately covered;

— to avoid silhouetting of subjects, for example where individuals enter a building from outdoors; where this cannot be avoided, or where strong shadows are present, consideration should be given to the use of supplementary (possibly infra-red) lighting;

— evenness of light distribution across the face as target subjects walk through the zone of recognition (see Figure 3);

— sufficient illumination of the faces to avoid motion blur (due to slow shutter speeds) or sensor noise impacting on the image quality.

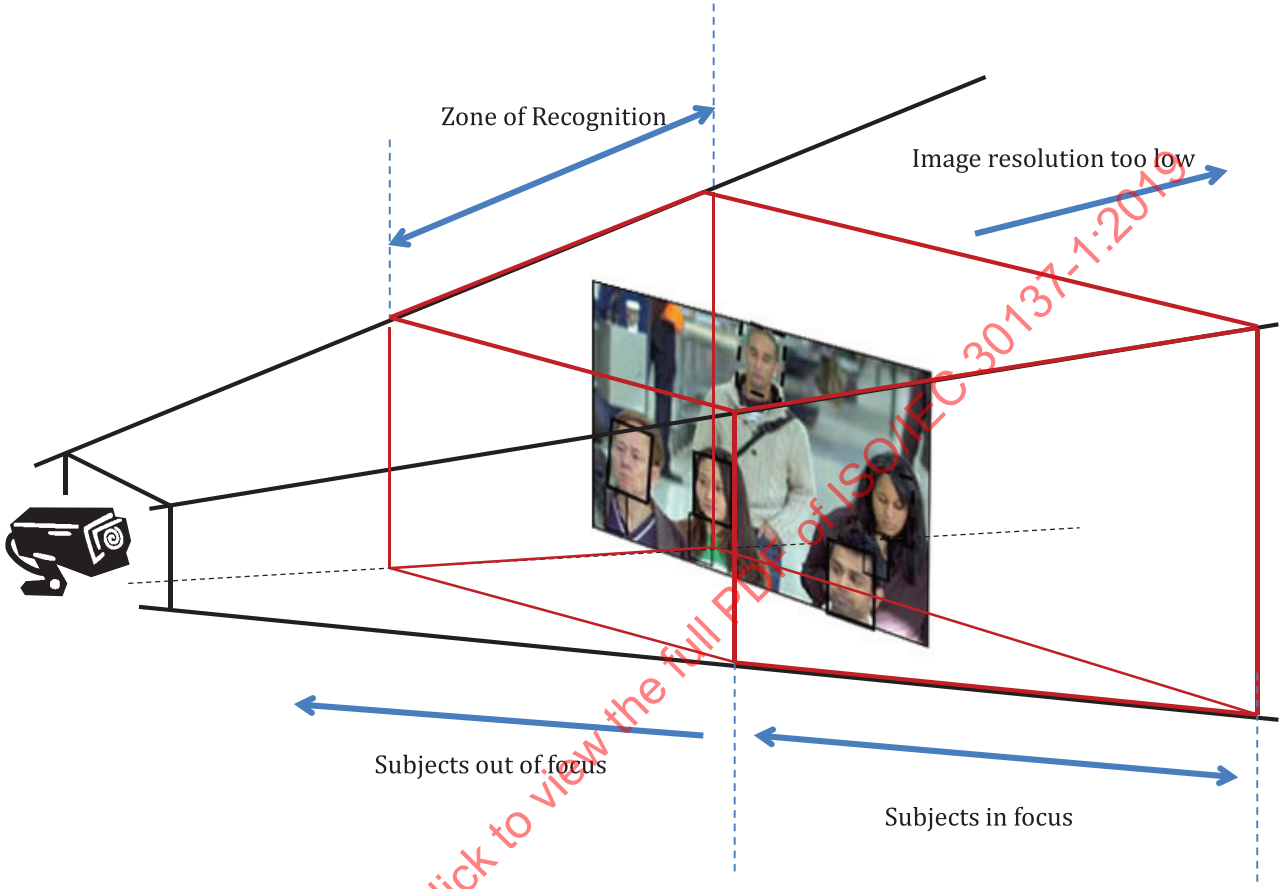
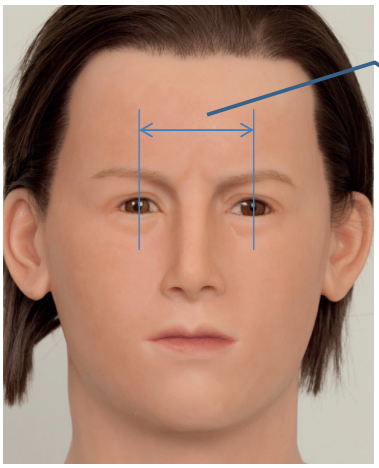


Figure 3 — Zone of recognition

Note that whilst IEC 62676-4 was not written with AFR applications in mind, it does nevertheless provide useful information on the field of view of the cameras in relation to object or target size for various scenarios, including crowd monitoring, detecting the presence of an individual, recognition of a known individual and identification of an unknown individual (see Table 1 for an explanation of these terms).

For AFR applications the camera's horizontal field of view in metres should be mapped against the horizontal image frame size in pixels, to obtain an effective resolution of pixels per metre. This can also be considered as pixels per millimetre when assessing likely biometric performance, based upon existing research which tests the inter-pupil distance or pixels between the eye centres. For AFR applications the pixels per metre at the furthest distance that subjects are expected to be captured in focus should be considered. For example a system may be optimised to work with facial images having approximately 95 pixels between the centres of the eyes, which is a known plateauing of performance for many current facial recognition systems[5][6]. Assuming a typical inter-eye distance (IED) for adults of 63 millimetres this will require a resolution of 1,5 pixels per millimetre (95/63) or 1 500 pixels per metre. To achieve this level of resolution with a high definition camera which has a frame size of 1 920 × 1 080 pixels, the horizontal number of pixels in the camera sensor (1920) is divided by required pixels per metre (1 500) resulting in a maximum horizontal capture width of 1,27 metres.

NOTE In practice the actual optical resolution of the VSS camera system is always lower than the value obtained using this method. Annex D contains more information on image resolution and methods for assessing camera performance.



A typical IED in adults is around 0,063 metres (63 millimetres)

[6] Dodgeson N,A (2004) - Variation and Extrema of Human Interpupillary Distance. SPIE, Proceedings Vol 5291.

A full HD camera has a resolution of 1920 by 1080 pixels

IED	Maximum horizontal capture width	Reference
20 pixels	6,04m= 0,063*1920/20	Minimum value examined in FIVE study
50 pixels	2,41m= 0,063*1920/50	Minimum value recommended in 8.5.1
80 pixels	1,51m= 0,063*1920/80	Maximum value examined in FIVE study
95 pixels	1,27m= 0,063*1920/95	Maximum value recommended in 8.5.2.1

Figure 4 — Relationship between the IED in pixels and the maximum capture width of the camera

7.5.2.2 Camera positioning to capture images for automated gait analysis

A gait recognition silhouette is the video frame image of a person represented as a solid shape of a single colour, usually black. The edges of the silhouette match the outline of the subject.

Gait analysis can be deployed for identification in multi-camera surveillance scenarios. View-point independent rectification is used to calculate (for example) side view coordinates for multi-camera video frames. Low frame-rate (1 fps to 5 fps) video recordings made with still image cameras or video surveillance systems are compatible with normalized gait silhouette sequences. In order to capture all the details of a gait signature, a minimum of one full gait cycle (i.e. two full steps) is required. Figure 5 illustrates the silhouettes of the phases of one full gait cycle.

a b c d e

Key

a), c), e) stance phases

b), d) swing phaces

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

Various automated methods have been developed for 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[10] view of a gait sequence. The subject should walk in a straight line perpendicular to the camera line of sight (lateral view) as illustrated in Figure 6. As an addition to this a similar recording showing frontal and back (dorsal) views may help in gait recognition when a person is followed through different scenes.

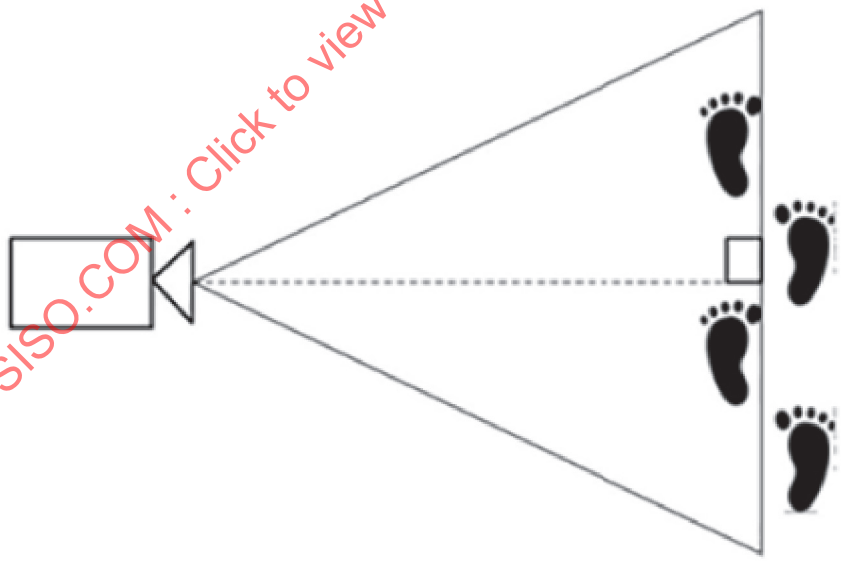


Figure 6 — Subject should be captured from the side view

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

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[11] 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.

Given the constraints in terms of camera placement and environment required for automated gait recognition, the applicability of this technique for subject recognition or tracking in real word scenarios is limited.

7.5.3 Infrastructure considerations

7.5.3.1 Power

Power consumption estimates should be made for

— cameras;

— network devices;

— digital video recorders or more general network storage devices;

— processing infrastructure.

The power consumption of such devices is typically published by manufacturers, such that power requirements can be estimated.

7.5.3.2 Network requirements and compression

Conventional VSS implementations transmit video imagery over a network for storage or processing or both. Typically, this is done on a sustained basis without interruption and the quantities of video data are large.

The bandwidth of the network should be sufficient to allow video data to be streamed on a sustained basis. The rate at which data flows over the network increases linearly with

— the number of cameras;

— the width and height of the camera images;

— the frame rate;

— the bit-rate of the compressed video.

Video data in VSS implementations is almost always compressed to conserve network bandwidth and storage space. Compression is usually implemented within the camera and is almost always lossy. This means that the original video from the sensor cannot exactly be recovered on decompression. The loss is minimized by

a) specifying a high bit rate, and

b) using modern compression techniques, such as those specified in ISO/IEC 14496-10.

Such compression techniques achieve good performance, in part, by including motion compensation. The bit rate is proportional to the network data flow rate. The bit rate is also a measure of video quality with low bit rates corresponding to high compression ratios and more loss of information. The recommended approach is to install sufficient network bandwidth to accommodate bit rates high enough to cause immaterial loss to the video data. The definition of "immaterial" should be based on automated recognition error rates, or on human judgements on the visibility of detailed features of the human face.

NOTE In NIST's FIVE study[12], using ceiling mounted high definition cameras producing 30 frames per second with pixel dimensions 1 920 × 1 080, the AVC codec was configured to produce a data rate of 24 megabits per second from each camera.

7.5.3.3 Storage and retention

When one or more cameras stream video data to a dedicated digital video recorder or other storage equipment on a continuous basis, it will ultimately fill the device. Therefore, it is necessary to establish a retention window sized to accommodate operational review objectives and cost constraints, in addition to complying with relevant governance policies and data protection legislation.

7.6 Biometric software

7.6.1 General

Regardless of the specific mode being used, the biometric subsystem performs a number of tasks:

1. detecting the presence of individuals within the field of view of the camera;
2. extracting the information required to create a biometric template for each individual using one or more frames from the video stream;
3. generating the biometric template(s);
4. comparing the template with those templates stored in a reference database and returning possible matches for review by the operator.

The following sections look at these in more detail from the perspective of an AFR application.

7.6.2 Face detection software

This is the first stage in the recognition process, before any matches can be made by subsequent processing. The speed and accuracy of the face detection algorithm are key factors in determining the overall performance of the system. Detecting faces in video is computationally expensive. It is especially challenging where this has to be done in real time and where there are multiple faces in the field of view of the camera, but where each individual is moving and may only be in the optimum position (the zone of recognition) for a matter of seconds.

The task is made harder as faces may be partially obscured by others, or the target subject may not be looking towards the camera and as a result may not be detected. At the other end of the scale, images and logos on clothing, or even static objects within the field of view can be misinterpreted as faces, resulting in the creation of templates that are of no value. Developers of such software set thresholds for various parameters which need to be met before the image is classified as a face. If these are set low (a "liberal" detection policy) to ensure that all "real" faces in the video stream are detected, it will usually lead to an increase in the number of spurious templates that are created. If they are set high (a "conservative" detection policy), while the templates generated will generally be of better quality, there is a risk that many target subjects will not be detected at all and thus will not be searched against the watchlist. Some algorithms may also set limits on the total number of faces that can be detected simultaneously (i.e. in any one frame) in order to avoid bottlenecks downstream.

A particular challenge in VSS applications is the frame rate at which the video cameras are operating. In most applications this is at least 10 fps and may be much higher, potentially generating many tens or hundreds of templates of each individual during the time they are in the field of view of the camera. Suppliers of face detection software may use various techniques (such as sampling only a subset of all available frames) in order to address this, but this may mean that the opportunity to create the "best" template of a target subject is missed.

The optimum settings for the face detection software will vary with the application. Time should be allocated during the commissioning of the system to determine the optimum balance between speed, accuracy and reliability with which faces are being detected in the operational environment.

Some face comparison software uses biometric feature sets created by combining features extracted from multiple face images from a single individual captured from different video frames. In this case multiple face images of one person are continuously acquired by the use of face tracking algorithms

(often different from face detection algorithms). Face images are then selected based on face quality metrics and similarity to previous acquisitions (to minimise the risk of face tracking errors) processed and combined into a single feature set.

7.6.3 Face comparison software

As with face detection software, the comparison software also varies in speed and accuracy, and a supplier may develop different versions which give preference to one or the other, and/or recommend specific processors and hardware in order to achieve optimum performance. A particular algorithm may be very suitable for post event searching of large databases where the probe and reference images have been collected under ideal conditions, but may perform very poorly in a real time environment where the probe images are from video surveillance cameras and captured in far from optimum conditions. Developers of AFR software may use a variety of techniques in an attempt to overcome problems with poor quality images, for example “averaging” faces across several video frames, or the use of 3D models to generate multiple templates from a single image. However, a detailed discussion of algorithm design is beyond the scope of this document.

7.6.4 Algorithm selection and testing

A supplier generally quotes figures for the performance of their algorithm, but it is important to bear in mind that these results may have been obtained under optimal conditions very different to the operational environment now being considered. Independent testing (e.g. such as that regularly conducted by NIST) may provide more reliable performance data but even this may not be representative of what will be obtained in practice. It is important that time is allocated for testing during the procurement and commissioning process to ensure that the system is delivering the anticipated level of performance and that it has been correctly configured for the specific environment and operational need. Guidance on specifying performance metrics can be found in 11.5.

Advice on such testing can be found in ISO/IEC 19795-2.

7.6.5 Other (non-biometric) software

There may also be supplementary software, not directly related to the biometric recognition function but required as part of wider business processes, including:

— management information software to monitor and manage the efficiency and effectiveness of the system;

— auditing software to record usage of the system and to support investigations, e.g. following a request for information by a member of the public or allegations of inappropriate use of the system;

— business continuity software to ensure continued availability of the system and or data in the event of power failures or system outages.

7.7 Computational requirements

7.7.1 General

The computing power required varies with the use case and the specific scenario. In the most demanding cases the system may consist of multiple high-resolution cameras, all running at high frame rates and each with multiple target subjects within the zone of recognition, coupled with an operational requirement for real time searching against a large watchlist of many thousands or tens of thousands of reference images. If the recognition time (see 11.5.2) is to be kept within acceptable limits, significant computational power may be required, possibly requiring dedicated rooms, fitted with suitable cooling systems, to house the hardware. In other less demanding scenarios a well specified laptop may be quite sufficient.

Within the types of systems covered in this document there are a number of places where lack of adequate processing power or capacity may result in computational performance bottlenecks:

— the communications network required to transmit the images from the cameras to the biometric subsystem;

— the software to detect and extract location information on each target subject within the zone of recognition of each camera;

— the software to create templates for searching against the watchlist;

— the biometric matching software to search and compare the templates of the probe images against those in the watchlist;

— the retrieval of potential matching images from the database along with any additional data which needs to be displayed to the operator for review.

7.7.2 Core biometric processes

Biometric recognition is broadly done in three phases. First is image processing for detection and tracking, followed by feature extraction, and finally searching against a watchlist. Regardless of biometric modality the cost of processing video sequences is larger than in still images, as the amount of data is greater. While some algorithms may elect to operate on only select subsets of the video data, there is cost in isolating the frames of interest. In general, the computation cost of the image processing tasks scales linearly with

— the size of the images;

— the duration of the video sequence;

— the number of people present in the video;

— the duration of their appearance.

Computation time varies greatly between algorithms, scaling linearly with

— how many (true and false) faces are reported;

— how many tracks are produced (and whether tracking integrity is lost such that an individual track is broken into several parts).

In general, the cost of the search is much smaller (depending on the size of the recognition template) but scales linearly with

— the number of templates extracted from the video;

— the number of enrolled templates (although sub-linear dependency is also known).

NOTE 1 In NIST's FIVE study[12] the time taken to process one second of a 1 920 × 1 080 video containing zero faces varied from about 0,4 s to 4 s for the most accurate algorithm and to 20 s for competitive alternatives.

NOTE 2 In NIST's FIVE study[12] the time taken to process clips of scenes containing up to 7 persons varied from 3 s to 11 s for the most accurate algorithm, and up to 80 s for other accurate suppliers. These figures are per second of 1 920 × 1 080 video.

NOTE 3 In NIST's FIVE study[12] most algorithms produced templates whose sizes grew linearly with the duration of the input video clip. Some algorithms, however, produced templates whose sizes were independent of the duration of the input. This might be achieved via best-frame selection, or by integrating information over time.

Given considerable variation in the various algorithm processing times, it is necessary to couple the hardware procurement specifications with those of the algorithm. It may not be sufficient to specify time limits in a procurement document, as this may have adverse accuracy consequences.

7.7.3 Reducing computational expense

There are a number of techniques which may be employed in an effort to reduce the computational power requirements, including:

— In-camera processing of the images can substantially reduce the volume of data which has to be transmitted over the network. For example it may be possible to detect and extract the facial image data from the video frames and only transmit the best images to the biometric subsystem for searching. However, in many scenarios there will be a requirement to view or record all of the data from the camera so this approach may not be acceptable, or may also require a separate video stream for viewing or recording the original data.

— Reducing the frame rate, applying compression or reducing the camera resolution can significantly reduce the volume of data to be transmitted over the network. However in each case some data is lost which may have a negative impact on other aspects of the performance, in particular accuracy.

— Detecting a human target subject moving within the field of view of a video camera in real time requires significant computational resources, especially if the frame rate is high and/or there are multiple target subjects within the frames. Some common techniques are described in Annex A. Possible solutions to reduce the computational overheads include only sampling a subset of all available frames, or raising the detection threshold, but the effect of both of these may be that some target subjects are missed [i.e. the failure to detect rate is increased. See 11.5.2 for more details].

— There may be an assumption that higher quality images, rich in detail, will produce better (and larger) templates, and that this in turn will lead to better search accuracy. However, the relationship is not that simple and the processing power and time taken to create a biometric template from the detected target subject(s) varies significantly with algorithm and supplier. As templates are proprietary and are closely coupled with the matching algorithm, selection of the most appropriate one for a given scenario should take into account a number of factors, including the speed with which a response is required, accuracy (given the type and quality of data, both of the probes and those in the watchlist), the size of the watchlist and the available computing power.

— Techniques for minimising the time taken to display responses to the operator include reducing the number of candidates returned (either by raising the matching threshold or limiting the length of the list) and reducing the size and resolution of the images initially displayed (e.g. higher resolution images are only provided if the operator wishes to take a closer look).

7.8 Specification for reference image database

7.8.1 General

Other clauses in this document give advice on the selection and positioning of VSS cameras to maximise the likelihood of capturing high quality probe images. However, the quantity and quality of reference images in the watchlist or database being searched (often referred to as the “gallery”) also plays a very large role in determining the performance of the system.

7.8.2 Reference database size

If a particular target subject's biometric data is not stored in the system then a “match” is not possible, no matter how good the quality of the probe images is. However, as the size of the gallery increases, so does the probability of the system returning false matches with comparison scores above the threshold. If this occurs frequently, operators may become disillusioned with the system and as a result may fail to recognise or act on a true match when one does occur. Large galleries may also result in multiple candidates being returned with the risk that even when a true match is present it may be pushed towards the bottom of the list and consequently not recognised as such by the operator.

For real time searching against watchlists, the number of target subjects whose templates are stored in the gallery normally needs to be limited to no more than a few thousand if acceptable performance is to be maintained. For non-real time applications (e.g. post event investigation following an incident) the

gallery size may vary from just a few (e.g. if the aim is to find a small number of specific target subjects captured on different cameras or in multiple locations) up to many millions.

NOTE In NIST's FIVE study[12] of video surveillance in a transport terminus, the percentage of target subjects in the watchlist that were not identified (the false negative rate) increased from 21 % to 35 % as the population size increased from 4 800 to 48 000 and the threshold was increased to keep the number of false positives constant.

As the size of the database increases, the use of techniques such as binning or filtering may be helpful in restricting the search space. Note that such techniques rely on metadata associated with the image and if this data is not present or is not reliable, the use of such techniques may actually reduce rather than improve search accuracy.

7.8.3 Reference image quality

Reference images stored in the watchlist should be of the highest possible quality. For facial images, ISO/IEC 19794-5 provides guidance on how to obtain suitable images, and ISO/IEC TR 29794-5 describes techniques that may be used to assess quality. Modern facial recognition algorithms can deliver very high levels of accuracy when searching such images against a database of similarly controlled images. However, in VSS applications the probe images coming from the video cameras rarely meet all of these criteria, and while it may be possible to control the lighting, the positioning of the cameras will typically result in the target subject's face being captured from slightly above or to one side. When they are available, the storing of multiple images of each target subject in the database is recommended to improve the performance of both the automated and human comparison processes. Such images do not necessarily need to all be of high quality (in the sense of being fully in accordance with best practices described in ISO/IEC 19794-5) to still be of value in a biometrically enabled VSS application; facial images captured under a range of lighting conditions, at different pose angles and with different cameras/lenses may still generate high comparison scores, especially where they are comparable to the conditions under which the probe image has been captured.

NOTE 1 In one study [FIVE] of non-real time video-based investigation, in a sports arena, accuracy from a single frontal photograph alone was compared with frontal plus the addition of non-frontal quarter-left and quarter-right photographs to the reference database. Accuracy gains varied by algorithm, resulting in 15 % and 60 % fewer misses at rank 20. More modest gains were also seen at rank 1.

NOTE 2 In the same study the inclusion of additional non-frontal reference images in the database proved of less value in real-time identification, where a high threshold is typically used. Identification misses reduced by around 10 % using the two additional non-frontal enrolments.

Regardless of the biometric mode that generated the potential match, it is in most cases the corresponding facial images that are used by the operator in reaching a decision. As explained in Clause 10 this is a difficult task, especially in a real-time application where a quick decision is required so that appropriate action can be taken. Images showing the target subject with a variety of appearances (different clothes, with and without facial hair, glasses or hats etc.) can all help the operator to arrive at the correct decision. If the images are of sufficient quality that a template can be created from them, they may be added to the searchable database. If a template cannot be created from an image it may be linked (via metadata) to images of the same target subject from which templates have been created. Making these available to the operator along with probe image and the image(s) returned with scores above the threshold may improve their performance.

7.8.4 Reference database maintenance

Database sizes tend to increase over time, as more images from more persons of interest are added. As outlined in 7.8.2 this can increase the utility of the system by expanding the population of interest but can lead to more false positives. The database growth rate equals the rate of addition minus the rate of deletion. It is common for images not to be deleted at all. However, the database owner should establish a systematic process to curate the contents of the database, by

— deleting templates from poor quality photos;

— replacing with templates from new improved quality photos;

— deleting templates older than some pre-determined number of years;

— deleting templates from persons now above some age limit;

— deleting templates from subjects according to known case dispositions (e.g. death, incarceration);

— reviewing criteria for risk.

8 Multiple camera operation

There are two principal scenarios to consider regarding the use of multiple cameras in a biometrically enabled VSS:

1. overlapping — where more than one camera is covering the same zone (i.e. the zone of recognition overlaps);

2. non-overlapping — where the cameras are covering completely different zones (i.e. there is no overlap in the zone of recognition).

When the Type 1 setup shown in Figure 2 cannot be used, an array of cameras mounted around the point of visual saliency and attraction (e.g. around a monitor displaying traveller information in airports) may provide a better chance of capturing a frontal image. Alternatively, cameras may be distributed around the environment, but all focused on a particular region in order to ensure that at any given time a target subject within that region is facing at least one of them.

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 necessary for any single 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.

Cameras with a wide field of view covering a large region may be supplemented by cameras with a narrower field of view covering a subset of that region, but which as a result are able to capture biometric samples at higher resolutions.

With any multiple camera setup the AFR's ability to manage target subject tracking becomes a key issue to avoid unnecessary additional (both false and positive) alarms once an operator decision has been made. Target subject tracking may also be exploited to then automatically provide a current location for the target subject to assist with responses.

Multiple cameras used by AFR systems should not remove the capability provided by manual or automatically controlled cameras (e.g. PTZs) to assist operators and responders in validating a potential match and to be able to see the target subject's current position, clothing, luggage, and any associates.

9 Interfaces to related software

In some scenarios, the biometrically enabled VSS may communicate with other systems, such as an access control application. Real-time communication between the VSS and an access control application is in most cases based on one of two protocols (Wiegand or OSDP) while non real-time communication (synchronising databases, updating activity and audit logs) is usually based on IP protocols.

10 Guidance for operator assistance

When an alert is generated by the biometric subsystem, in most cases there is a requirement for a human operator to assess the alert against a record held in the watchlist and to make a decision on the most appropriate action to take. Regardless of the biometric mode that triggered the alert the operator's decision is generally based on a comparison of the target subject's facial image with one or more images on the watchlist.

However, an operator's ability to do this accurately when presented with facial images which are not familiar to them is in general very poor, even for those with many years of experience. Recent research[15][16] indicates that the cognitive ability to perform this task is distributed unevenly across the population, with some individuals (prosopagnosics) having exceptionally poor competence, while a small proportion of individuals demonstrate a high level of competence at this task.

The careful selection and appropriate training of operators therefore plays a key role in determining the overall (end-to-end) operational performance of the system. While training courses in facial comparison are available, this is still an area of ongoing research and trainers and examiners should be carefully assessed against current practice.

The operator's workstation should be configured to enable them to work effectively, using tools and procedures developed specifically for the purpose and taking account of the time available to make a decision. There may also be merit in having more than one individual review the images to arrive at a consensus decision, or to have the operator perform just an initial screening of the output from the biometric subsystem and to then pass likely matches on to a second (more highly trained) individual for more detailed analysis.

The pictures of target subjects detected by the system in real time, and that result in one or more candidates above the comparison threshold, should be presented on the screen next to those image(s) returned by the biometric comparison process. This way the operator will have the ability to visually verify the output from the biometric subsystem and to confirm or reject the recognition. If they are available it can be helpful to display multiple images of the same target subject from the watchlist, taken under different conditions and at different times. Likewise, the ability to view video clips (if available) rather than just still images may also be beneficial to the operator.

In many applications the number of false alerts generated by the system is significantly higher than the number of correct matches, and the time available to make a decision is very short. There may also be multiple potential matches returned above the system threshold and processes should be in place to handle such situations. Where it is available, additional information from the system (and from appropriate decision support software) may be useful in helping the operator reach the correct decision.

As well as taking care when selecting suitable operators, ongoing assessment of their performance is recommended, e.g. by holding regular training exercises in an operational environment and lessons learned activities. In scenarios where very few true matches are expected, a mechanism for injecting "matches" into the workflow may be useful to monitor and ensure the vigilance of operators.

11 System design considerations

11.1 General

The following clauses provide information relevant to the design and implementation of a biometric system for use with video surveillance cameras.

As is the case throughout this document, the assumption is that in most cases this will make use of facial recognition technology, although the guidance can also be adapted for other modalities.

11.2 Establishing the business requirements

Prior to starting any design activities, it is first necessary to understand the business requirements that the system is intended to address, including how the biometric components fit within the wider operational processes. For example:

— Is the use of biometrics really the most appropriate solution or are there other approaches that could achieve similar results?

— If a biometric solution is to be implemented, what mode is most suitable and how will it be used (e.g. real time identification or post event analysis)?

— How will existing business processes (both internal and external) be impacted by the introduction of the new system?

— How will the results output by the biometric subsystem be used in normal operations, will they be presented to an operator for a quick review or will they require more detailed examination?

— What training will the operator require and what action will subsequently be taken as a result of their decisions?

— What are the legal and societal implications of introducing a biometric capability in this particular environment?

— Is it possible that the system results and operator decisions may eventually need to be presented and justified in a court of law?

11.3 Site survey

Unless working with data supplied by third parties over which there is little or no control the first step when planning for the implementation of a biometric capability for use with video surveillance cameras is to undertake a detailed site survey of the location where it is to be deployed. This may initially be conducted by the customer, but for integration of FR into fixed infrastructure the site survey should also be undertaken by the supplier of the biometric software and any other third parties such as a systems integrator. An assessment of the numbers of individuals typically present at different times of the day (and night), the way they move through the environment, the level, quality and variability of the lighting, location of existing cameras, opportunities for placing additional cameras etc. all contribute to an understanding of the challenges and opportunities that the specific location provides. A site survey can be an opportunity for the customer to assess the level of knowledge and expertise of potential suppliers, but also for suppliers to improve their understanding of the user requirements and, if appropriate, to manage user expectations regarding what is realistically possible with current technology given any other constraints that might apply (e.g. positioning of cameras, lighting, overall cost).

As a minimum the site survey should seek to elicit answers to the following:

— Stakeholders

  — Establishing who owns, maintains, operates or otherwise has a responsibility for all aspects of the environment or premises where the system will be deployed.

  — Are there any dependencies on external systems or stakeholders that may send or receive data, or that may otherwise be impacted by proposed changes to the existing processes?

  — Have other interested parties such as regulators, representatives of civic society and organisations representing employees been consulted?

— Cameras

  — What cameras are currently in use [analogue/digital, frame size, frame rate, focal length, fixed or moveable (e.g. PTZ cameras)]?

  — Are they suitable for use with the biometric modalities being considered?

  - Do they need to be upgraded, replaced or supplemented?
- Environment
  - How many individuals are likely to be present within the environment covered by the cameras throughout the day/week/year (min, mean, max)?
  - What are the most common routes (including direction of travel) that individuals take as they pass through?
  - How long do they typically spend in the clear view of each of the cameras?
  - Are there any existing choke points that are also particularly well suited to facilitate the capture of good quality biometric data?
  - Is there an opportunity to introduce choke points or otherwise modify the flow of individuals?
  - Are there any attractors or distractions that may affect the target subjects' behaviour (for better or worse)?
  - Is it possible to introduce attractors at strategic locations within the environment?
- Lighting
  - Is lighting primarily artificial (i.e. controlled) or natural (i.e. uncontrolled and likely to vary throughout the day and with the weather)?
  - What type of lighting is used and how bright is it?
  - Is the lighting level consistent or are there particularly bright or dark areas that might cast shadows?
  - Is there an opportunity to introduce additional lighting?
- Network infrastructure
  - What type of communications network is currently installed for the cameras?
  - What standards are currently in use (e.g. video codecs, other data formats, interfaces to external systems)?
  - Does the network have sufficient capacity/speed? Will it need to be upgraded?

11.4 Size and content of the watchlist

Subclause 7.8 provides guidance on the selection of images for the watchlist, and how its size impacts on the performance. In designing the system it is important to be clear what the watchlist will contain and how the data in it will be used by the biometric sub-system. For example:

- Where will the images in the database come from and what is the overall quality and quality distribution?
- If they are facial images, are they passport style photographs, in accordance with ISO/IEC 19794-5, facial images extracted from CCTV footage or similar uncontrolled environments, or a mixture?
- Will there be multiple images/templates per target subject? How will these be linked?
- What quality criteria will be applied when determining if an image is of sufficient quality to be enrolled; how will the optimum enrolment threshold be determined?
- What metadata is available and how will it be stored and used by the system?

— Is there a requirement to be able to filter searches or logically partition (bin) the data based on parameters such as gender, age, ethnicity or other parameters?

— What requirements and processes are there for creating, reading, updating and deleting records from the database?

— How many templates will be stored in the database?

— Where will the database be physically located?

11.5 Performance requirements

11.5.1 General

The term "performance" in biometric systems is often used to refer solely to "accuracy" but this is only one of many factors that contribute to determining how well an operational biometric system is working. In Figure 1 (system architecture), each of the components shown has an impact on the overall performance of the system, be it the environment, the quality of the video cameras, network bandwidth, accuracy of the face detection and comparison algorithms, the quality and number of images in the watchlist, or the way in which the operator interacts with the system and is able to make a correct decision when a potential match is returned. When setting performance requirements, it is therefore important to consider the end-to-end performance of the whole system and not any one component in isolation.

The performance requirements for a real time application (where speed of response is critical) are generally very different to those used for post event searching, where the priority will typically be to maximise search accuracy, and in particular to minimise the number of false non-matches (even though this may be at the expense of a higher false match rate).

11.5.2 Key metrics of performance

Performance requirements for the end-to-end service should be identified as a part of the user requirements capture process. These then need to be translated into measurable system metrics. Note that in most applications the end-to-end service also includes one or more operators whose role is to review and confirm or reject potential matches returned by the biometric subsystem. This human element can have a significant impact on the overall performance of the system but is often excluded from system performance metrics due to the difficulty in measuring it. See Clause 11 for further information on the role of the operator.

Key performance metrics include:

— failure to detect rate: the proportion of individuals whose faces appear in the field of view of one or more cameras but which are not detected or accepted for subsequent processing and comparison;

— failure to acquire rate: the proportion of individuals whose faces appear in the field of view of one or more cameras but which fail to result in the creation of a probe or reference template for subsequent facial comparison;

— true recognition rate: the proportion of individuals whose faces appear in the field of view of one or more cameras, and that are also in the watch list, and which are detected and matched correctly. Note that in some scenarios it can be difficult to determine this figure as the correctness of an alert might not be established in all cases.

— false alert rate: the proportion of individuals whose faces are detected but which result in a false match to a different individual in the watchlist. In some scenarios the number of false alerts may be substantially larger than the number of true recognitions and this is a key factor in determining the workload on the operator;

— recognition time: the average time from when a face appears in the field of view of the camera(s) to the time when the facial recognition system declares a potential match. Note that in most

applications the actual operational recognition time is longer as it will also include additional time taken for the operator to review and confirm the match.

NOTE The definitions above refer to the faces of individuals appearing in the field of view of cameras. However, it is also possible to use appearance of the face within the zone of recognition as the relevant criterion.

11.5.3 Presentation Attack Detection (PAD) performance metrics

Some biometric systems include additional PAD software to detect presentation attacks or other attempts to subvert the system. The use of such software is likely to have an impact on the overall performance metrics (e.g. as a result of incorrectly classifying a legitimate presentation as a subversive one and vice versa). Therefore, it may be necessary to specify performance metrics for the system both with and without such PAD software running, and/or specify performance metrics for the PAD subsystem. Examples include:

— Attack Presentation Classification Error Rate (APCER): the proportion of presentation attacks incorrectly classified as normal presentations;

— Normal Presentation Classification Error Rate (NPCER): the proportion of normal presentations incorrectly classified as presentation attacks;

— Attack Presentation Non Response Rate (APNRR): the proportion of presentation attacks in which the PAD subsystem correctly fails to respond;

— Normal Presentation Non Response Rate (NPNRR): the proportion of normal presentations in which the PAD subsystem incorrectly fails to respond.

Further information relating to the performance and evaluation of PAD subsystems can be found in ISO/IEC 30107 (all parts).

11.6 Image data and metadata considerations

Both video cameras and digital still cameras are used for video enrolment and video surveillance. Image orientation is typically not a problem for VSS. Both JPEG and MPEG-4 Part 14 format definitions include metadata for camera parameters and camera orientation. JPEG uses EXchangeable Image File (EXIF) format as metadata and MPEG-4 Part 14 uses Extensible Metadata Platform as the metadata format. It is recommended to use these two formats for video and still image surveillance.

Annex A
(informative)

Other related (but non-biometric) video analytic techniques and applications

A.1 General

Automated tracking of subjects within and between cameras can play an important role in VSS but is not in itself a biometric capability or use case. Similarly, estimation of crowd densities through the use of video surveillance requires the system to be able to differentiate between individuals, possibly by locating and temporarily storing facial images/templates.

Such applications are not the primary focus of this document, but because a biometrically enabled VSS may also include this type of functionality, an overview of them is provided in this informative annex for completeness.

Video analytics consists of computer-assisted reproduction of the analysis that a human operator would do looking at the video footage from the surveillance cameras. The analytics software processes video to automatically detect people and events of interest for security purposes. Only once they have been detected can the individuals be identified, monitored and located.

While many video analytics techniques are not directly applicable to biometrics applications, the ability to recognise human beings in a video stream (and in most cases to locate and extract faces) is a pre-requisite for the subsequent biometric matching process. Hence this annex provides background information on techniques commonly used to detect, classify and extract information from video.

A.2 Real time alerts

Most alert conditions are defined by the intelligent video surveillance system user. They may be generic alerts, such as detecting an object or an item in the scene moving over a set speed limit. To trigger these alerts, only the properties of the object movements are analysed by the system. More specific alerts may be issued after the objects or their movement have been classified (e.g. discrimination between the passage of a human or an animal in an outside area). Related alerts based on conformity or non-conformity with a behaviour model entered in the system (e.g. an individual trying to open more than one car in a parking lot) constitute pre-defined alerts.

A.3 Video search for investigative purposes

Analytics processing makes it possible to index video content based on characteristics such as the shape of a person, their size, appearance, trajectory, type, as well as their model of activity. Stored as metadata, this information makes it possible to conduct spatiotemporal searches such as “find all footage with a person dressed in red passing in front of a certain building between two given dates”.

Video searches for investigations are executed both at the pixel and object level to achieve the scale.

They are grouped according to the following tasks:

— detection of changes;

— segmentation of moving humans;

— monitoring of humans;

— classification and identification of humans;

— classification of activities and behaviours.

A.4 Detection and segmentation

Detecting changes in video footage does not specifically target the movement of humans, but may highlight an image modulation. In order to segment moving humans it is necessary to be able to discriminate between those fluctuations in pixel values corresponding to consistent movements and those fluctuations caused by environmental changes.

EXAMPLE The system is intended to detect an activity in a scene under surveillance, in particular the movement of objects. It can also reveal the appearance or disappearance of an object (e.g. abandoned or stolen object). It is also used to automatically report accidental or intentional alterations in a camera: obstructions (dust, spider webs, moisture, paint and stickers), re-orientation and blur.

Several movement segmentation techniques have been proposed:

— Subtraction of the background

A first category of techniques consists of comparing each frame in a sequence to a reference image, called the background, which represents the undisturbed scene. The areas of change are formed of pixels with a difference in intensity that is above a threshold. Pixel-by-pixel subtraction between two images is very sensitive to the slightest environmental change, such as changes in lighting and movements inherent to a scene (e.g. the foliage of a tree blowing in the wind). In order to offset this problem, certain techniques continually adapt the background model to intrinsic changes in the environment. Subtraction of the background is a method that is particularly suited to indoor environments, where lighting conditions are controlled and where there is little activity (e.g. monitoring a hallway).

— Time-based difference

A second class of methods for detecting change is based on a time difference between a few consecutive frames. These frames adapt to variations in the time of the environment. On the other hand, they tend to be oversensitive to certain variations related to the movement of objects in the scene, especially if they move slowly. They often produce holes in the objects detected. These techniques therefore require smoothing treatment, with morphological operators and filtering of holes and shapes that are too small. In order to retain only significant movements and eliminate occasional movements, certain techniques draw up a map of the regions with a high level of activity, based on a movement pattern.

— Optical flow

Methods that analyse optical flow help to detect consistent directions of pixel change associated with the movement of objects in the scene. However, they require complex calculations that are difficult to do in real time. Optical flow is also sensitive to image noise.

A.5 Tracking

Many tracking techniques are based on mathematical methods that make it possible to predict a person's position in a frame based on their movement in the previous frames.

Tracking several individuals at the same time poses many challenges, including associating each person detected in a frame with the corresponding person in the subsequent frame. This matching is done based on the individuals' outlines, their characteristics (e.g. corners, area, ratios, etc.), or their model of appearance.

NOTE Occlusions (regions hidden by objects or other individuals) represent a major difficulty for tracking humans. A VSS can lose track of a target subject if they are totally or partially obstructed over a certain period of time. It can also be difficult to separate two individuals when they are very close or when one obscures the other.

A.6 Classification and identification of objects

Objects detected by a VSS are usually classified into different categories: human, vehicle, animal, etc. This classification may be done prior to tracking in order to retain only the trajectories of objects that are relevant for surveillance purposes.

EXAMPLE A human is usually presented as a form that is taller than it is wide, whereas an automobile would be wider than it is tall. Human gait has specific features, in particular, a certain periodicity. Therefore, systems recognize the nature of an entity detected based on its shape attributes and movement properties in general.

Object identification pushes recognition further; in addition to determining the class to which an object belongs, an identifier is also assigned. With surveillance, and in particular for access control or when searching for a suspect, the goal is to recognize a specific individual or decipher a particular vehicle license plate.

There are many environmental factors which impact on the system's ability to correctly analyse an image: bad weather, headlight brightness, dirty or damaged licence plate. In order to read a car licence plate, a system first locates the rectangle of the plate among all of the image's details. It then proceeds with optical character recognition. A licence plate filmed at an angle distorts the characters in the image and complicates the recognition process. In order to maximize the system's efficiency, plate recognition is done most often using specialized systems that concentrate on camera positioning and lighting quality.

A.7 Classification of activities and behaviours

Analysing and interpreting behaviours means recognizing movement patterns and extracting from them, at a higher level, a description of the actions and interactions. As is the case with all classification issues, a sequence of characteristics observed has to be associated with a model sequence representing a specific behaviour. The problem therefore consists of modelling typical behaviours, by learning or by definition, and finding a comparison method that tolerates slight variations.

EXAMPLE Hidden Markov Models (HMM), neural networks and Bayesian networks are among the most used techniques for modelling normal behaviour and thus for detecting abnormal behaviour. These techniques trigger an alarm based on statistical discrepancy with the inferred model of the scene. Predefined event detection methods also exist. These are based on a system of rules, such as triggering an alarm if an object bigger than a threshold value remains stationary for a certain period of time in a given region.

A.8 Crowd analysis

It can be important to understand a crowd's movements for security purposes.

EXAMPLE The crowd's trajectory and flow are analysed. When modelling the crowd and its behaviours, certain researchers consider the crowd as a whole and interpret the movements of the different parts. Techniques such as optical flow and HMM are used to model movements. Some models combine microscopic (individual) and macroscopic (crowd) analysis.

Annex B
(informative)

Societal considerations and governance processes

This Annex provides specific additional guidance and best practice advice based upon initial pilots of systems and lessons learned from other deployments of biometric technologies.

General advice in respect of non-technology issues in the use of biometrics for surveillance can be found in ISO/IEC TR 24714-1.

For organisations intending to deploy such a system, best practice recommends that a systematic approach be employed to ensure that all relevant issues are addressed in a timely fashion. These should be defined in a policy at the start of planning a system, and should cover legal issues (for example, addressing personal data protection, health and safety, and evidential use in courts of law), security considerations (including alerts to attempts at evasion), usability aspects, frequency and details of testing, etc.

A governance structure which brings together stakeholders and is put in place at an early stage in the planning process could assist in the development of a system that gains acceptance across the community it aims to serve.

Although the design and deployment of a legally permitted service is paramount, other issues of concern to the community can be captured through the creation of an ethics committee at arm's length from the operational authority and other governance bodies for the project. Such a committee could work within a previously determined ethical framework, that may be modelled on the principles which have been established in other technology areas or use the insights of research projects such as those in the Framework research programmes of the European Union: for example, RISE[17], HIDE[18], Tabula Rasa[19] and PRESCIENT[20]. By taking a wider perspective than just the protection of personal data and privacy, an ethics committee can consult groups who may not be usually represented (for example, the marginalised, transient visitors to the area under surveillance and those who are disabled), enquiring whether "just because it is not forbidden, it should be necessarily permitted".

An ethical principle which is often not considered is equal access to expertise about the operation of biometric systems. As described in this document, a biometric surveillance system is complex, with the interrelationships between components, interfaces and methods of assessing its performance known to the operating authority, but understood in depth only by those responsible for systems integration. In contrast, those individuals and groups which may be affected by the introduction of a surveillance system are unlikely to be aware of possible impacts on them. The ethics committee can reach out to technical departments in universities, colleges of art and design, etc. to find ways of representing contentious issues in more accessible terms. Resources which have been developed can then be shared with other projects.

For instance, the issue of proportionality of a proposed implementation may appear easy to assess. Whether it is designed to identify premium customers in a shopping mall, to exclude those who have been legally banned from entering areas under surveillance or to self-exclude from bars, casinos and betting shops, the deployed system could be accepted by the local community as a standalone service. However, by later connecting it as part of a network of systems across a city, thereby reducing monitoring costs and using it for general law enforcement purposes, the initial justification to the community may no longer be viewed as proportionate or acceptable.

Assumptions about the way biometric surveillance systems operate can lead to misunderstandings and sometimes to adverse consequences for the individual or for the effective operation of the system. For example, the non-uniform distribution of false alerts amongst the population (the "biometric zoo" effect) can result in certain individuals being repeatedly misidentified as persons on a watchlist, and others being particularly difficult to identify. In poorly designed systems, it may be relatively easy for

target subjects to evade recognition through the use of simple measures (e.g. by looking away from the cameras), and for others to be highlighted on suspicion of trying to take advantage of such measures; in either case, it will compromise the effectiveness of the system. Repeated stories in the media of such misidentifications could lead to a loss of trust in the technology.

During the design, deployment and testing of systems, there is a need for system operators to consider the collection and use of metadata, the management of watchlists and the possible evidential use of the identification of individuals in courts.

Metadata associated with the sensors and cameras, possible matches using the comparison software, the human operator resolution of alerts, etc. can either be discarded or retained for subsequent use in monitoring the health of the system, for research into improved performance or to support evidential use in courts. The exploitation of metadata whether for routine use or to allow access under exceptional circumstances requires careful analysis to ensure that excessive personal data is not retained, that data quality issues are addressed, and that search and use are facilitated. For countries with privacy and personal data protection laws, subject access requests may be permitted routinely, in which case, the metadata should be easily searchable for this purpose.

The operator should develop a clear policy regarding the management of watchlists. Although the suppliers of some commercial systems may claim that they can work with poor quality images, there is evidence that the recognition performance is degraded. If facial images obtained from video cameras are to be added to the watchlist, tests should be undertaken beforehand to demonstrate that the system will operate effectively, and that standards are developed which set a minimum quality threshold before images are added to the watchlist. In certain installations, more than one operator may use the video cameras with separate watchlists, necessitating a policy that minimises the potential for conflicting actions following a possible identification.

It is also important that the trade-off between usability and security issues is taken into account. ISO/IEC TR 29156, while not specifically intended for use with surveillance systems, does nevertheless define a methodology on how to balance usability and security requirements.

Where imagery and data from biometric surveillance systems is likely to be used in courts, the extraction and secure management of material, metadata and conclusions of operators needs careful consideration. Periodic tests should be carried out to demonstrate that the data continues to be accessible and processes remain fit for the purpose. Standards for training of operators monitoring the alerts from the automated system should also be developed. Confirmed instances of misidentification (or missed identification) provide a valuable resource for fine tuning the operation of systems.

Although data retention policies determine the period of time for which imagery and data is kept, consideration is needed for exceptional cases, for example where more data, collected over a longer period of time, etc. is required for research and testing purposes. Operational policies should allow for this, while minimising the opportunities for the abuse.

VSSs which have a biometric surveillance capability should, in general, be identified as using this additional functionality. Signage to alert to the deployment of overt systems should be placed in positions which will be visible to persons entering an area under surveillance, alongside indications of how to obtain further information. Work is underway to develop a standard for a visual representation of facial recognition, ISO/IEC 24779-5.

Annex C
(informative)

Case study: The use of AFR with VSS for traveller triaging at the border

C.1 General

In this example there is high volume of individuals (travellers arriving at the border) who need to be quickly assessed for risk on arrival at the border control check point. To keep queues to a manageable length, travellers are to be referred for further questioning (secondary inspection) only if the suspected risk is high.

The officer at the primary inspection point has some flexibility, within the terms of the agreed Standard Operational Procedures (SOP), in choosing if and when to ask additional questions of the target subject (and how many) with the objective of keeping the total wait time at the primary inspection lane within the defined service level (e.g. less than 30 s for 80 % of travellers). Those travellers that give cause for concern (for whatever reason) are referred to a secondary inspection point for further questioning.

To help the border control officer reach a decision regarding the appropriate course of action for each traveller, an AFR system is employed, linked to a biometric watchlist populated with individuals that are to be sent for additional processing.

Cameras are positioned such that it is possible to obtain high quality video footage of the target subject's face while they are standing at the primary inspection desk and this is then used to search the watchlist.

The biometric subsystem is configured with two comparison thresholds and the AFR search will result in one of three responses:

— A green "flag" from the system means that no potential match scoring above the lower threshold was found in the watchlist and, at least with regards to this aspect of the border check, there is no reason to send the target subject for additional secondary screening. Note that the officer may still decide to refer someone to secondary inspection for other reasons, not related to the AFR result.

— A yellow "flag" from the AFR system means that a potential match with a score somewhere between the lower and upper thresholds was returned. (This is sometimes referred to as a 'low or moderate confidence' match). This information is used to prompt the officer to ask additional questions to the traveller within the terms of the SOPs (i.e. without consideration of AFR results).

— A red "flag" from the AFR system means that a potential match scoring above the upper threshold was returned (a 'high confidence' match) and this results in immediate referral to secondary inspection.

C.2 Subjects

In this particular example the subjects are all travellers (without exception) passing though the checkpoint/document control.

C.3 Target subjects

The target subjects are those individuals whose biometric data is held in the watchlist(s) maintained by the border control operators.

C.4 Role of the biometric subsystem

The role of the AFR is to provide a quick indication to the officers at the primary inspection point if a traveller looks similar to someone in the watchlist, without requiring the officer to spend additional time to further manually examine his or her resemblance.

The phrase “looks similar” as opposed to “match” is very important as a lexicon for border officers. Similarity does not imply any negative connotation and should never result in inconvenience to travellers based solely on the fact that they resemble someone else.

C.5 Operator role

The operator in this example is the border officer sitting in the primary inspection booth. The officer will have been shown the current list of “wanted” individuals (as part of their normal shift handover procedure) but is not generally trained to be a face expert.

Their role here is to ask additional questions if prompted to do so as a result of the AFR system returning a yellow flag. They will then be sent either to the exit if the officer is satisfied the target subject is a bona-fide traveller, or to secondary inspection if they still have concerns. A red flag triggers an automatic referral to secondary inspection.

C.6 Performance objectives

In order to keep queue lengths to a minimum, the number of false referrals to secondary inspection should be minimized, with the vast majority of travellers being processed at the primary inspection point.

In this particular example, the “red alert” flag has been set such that the false alert rate is no higher than the value used for random referrals (e.g. <0,2 %). In other words 1 in every 500 travellers will be randomly selected for secondary inspection, regardless of the result of the AFR system. The use of the AFR system should not result in more than an equivalent number being referred to secondary inspection.

The selection of such parameters clearly requires a risk balanced trade-off between security and convenience/queue length. Choosing lower matching thresholds on the AFR system may result in more travellers on the watchlist being identified but would also result in a much larger number of yellow and red flags being returned and a corresponding increase in the number of legitimate travellers being referred to secondary inspection, with all of the associated inconvenience and delays, as well as the additional processing overheads incurred by the border control authority.

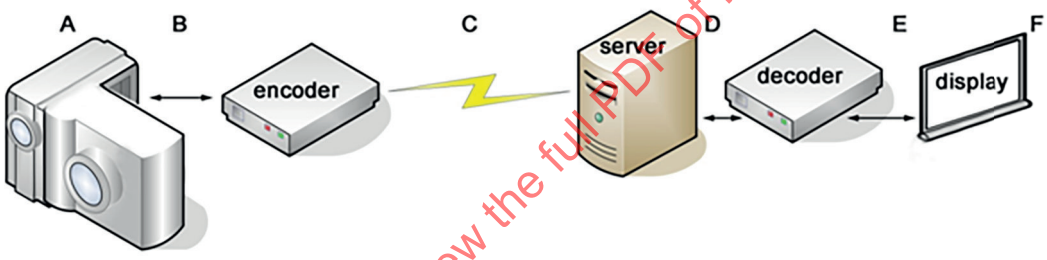
Annex D
(informative)

Video acquisition measurements

D.1 General

The measurements described in this annex are suitable for installation adjustment, calibration and for maintenance purposes to maintain high image fidelity. Adjustment can then be made to set the camera and illumination so that images taken are conformant with the relevant specifications.

The video camera performance measurements described here 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 video frame images on a computer.



Key

A camera sensor
B output
C transmission
D storage
E decoding
F biometric recognition software

Figure D.1 — Access points for video image measurements

The original video stream from the camera sensor is processed for output and encoded for transmission, often inside the camera body. After the storage a decoding step follows before the video is displayed or sent for processing by the biometric recognition software.

The H.264 [21] and MPEG-2[22] standards define different video encoding (video compression) formats and transmission schemes. For this reason the video transmission path may look different to the one shown in Figure D.1.

Access points (B) and (E) shown in Figure D.1 should be used for measurements.

In VSS 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 deforms the image and may be observed

as straight lines in the test target being rendered as curved lines in the image from the camera. Lens distortion correction post-processing is possible but lowers the resolution of the image.

D.1.1 Resolution

Resolution is one of the most important image quality factors. Resolution is a single frequency parameter that indicates whether the output signal contains a minimum level of detail information for visual detection. In other words, resolution is the highest spatial frequency that a video camera can usefully capture under cited conditions. The term resolution is often incorrectly interpreted as the number of addressable photo elements (the number of pixels in the sensor). Qualitatively, resolution is the ability of a camera to optically capture finely spaced detail. Noise, low dynamic range, strong vignetting and poor colour quality impair pattern recognition software's ability to resolve the details in an image. Lighting condition improvements are vital in resolving problems related to these factors.

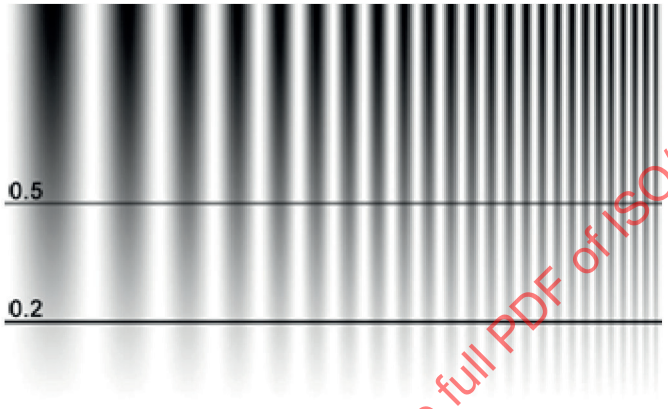


Figure D.2 — Variable frequency sine wave chart showing the SFR measurement principle

SFR is a multi-valued metric that measures contrast loss as a function of spatial frequency. Generally, contrast decreases as a function of spatial frequency to a level where detail is no longer visually resolved. This limiting frequency value is the resolution of the camera (see Figure D.2). Contrast values are marked showing the 50 % (0,5) and 20 % (0,2) levels.

The traditional method of measuring sharpness uses a resolution test chart. The measurement process starts with taking a video and capturing a still image frame of a resolution test chart containing bar patterns (see Figure D.3). Next, the captured image is examined to determine the finest bar pattern that is discernable as black-and-white lines. Finally, measurements of the horizontal and vertical resolution are made by using bars orientated in the vertical and horizontal directions, respectively. This procedure presents problems because it is manual and its results have a strong dependence on the observer's perception.

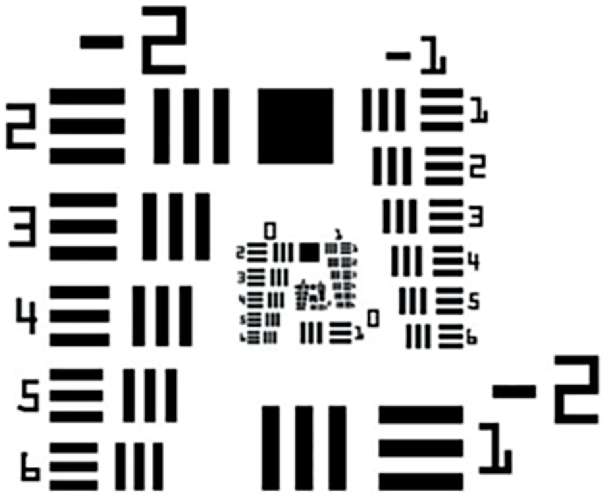


Figure D.3 — USAF 1951 resolution test chart

Furthermore, it delivers resolution results that correlate poorly with perceived sharpness since H.264 (MPEG4 Part 10)[21] or similar video compressions degrade video quality. Therefore, it is recommended to measure the MTF of the video camera system.

ISO 12233 contains a powerful technique for measuring MTF from a simple, slanted-edge target image that is present in the ISO 12233 resolution test chart. If video frames are captured from compressed video file then the MTF measurement shows the resolution of a compressed image and not the original camera image.

NOTE Modulation transfer function (MTF) is the scientific means of evaluating the fundamental spatial resolution performance of an imaging system, or components of that system. The spatial frequency response (SFR), analogous to the MTF of an optical imaging system, is one of four measurements for analysis of spatial resolution defined in ISO 12233 and provides a complete profile of the spatial response of digital still-picture cameras. In other words, MTF is the name given by optical engineers to Spatial Frequency Response (SFR). The more extended the MTF response, the sharper the image.

D.2 Measurements

D.2.1 Exposure metering

Standard Lighting Intensity (SLI) is approximately 200 lux to 500 lux (a lux is equal to one lumen per square meter) with ±10 % uniformity over the test target area. It is recommended to have a higher video light intensity on the subject in order to compensate for the ambient light. It is particularly important to ensure good environmental conditions because poor conditions usually result in the creation of low-quality biometric references, which lead to poor performance through increased biometric recognition error rates. Strong uncontrolled ambient lighting in the vicinity of the subject or camera should be assessed to determine if it will deliver the level of security required whilst minimizing possible interference to the recording of the subject as a result of excessive or uneven illumination.

Exposure measurements may be done by taking a picture with the video camera of a grey background or grey target and measuring the resulting pixel RGB values with an image processing application.

D.2.2 Standard test chart setup

Use the standard test charts shown in this annex to measure resolution, noise, dynamic range (indirect method) and colour accuracy (and white balance). Resolution is usually well above requirement limits when digital cameras are in use and lights are set properly. Resolution measurement helps to determine