



Video surveillance standardisation activities, process and roadmap

*ERNICIP Thematic Group
Video Surveillance for
Security of Critical
Infrastructure*

James Ferryman, Ph.D.
University of Reading, UK

August 2016

The research leading to these results has received funding from the European Union as part of the European Reference Network for Critical Infrastructure Protection project.

Video surveillance standardisation activities, process and roadmap

This publication is a Technical report by the Joint Research Centre (JRC), the European Commission's in-house science service. It aims to provide evidence-based scientific support to the European policymaking process. The scientific output expressed does not imply a policy position of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use which might be made of this publication.

JRC Science Hub

<https://ec.europa.eu/jrc>

JRC103650

ISBN 978-92-79-63952-4

doi:10.2788/92267

© European Union, 2016

Reproduction is authorised provided the source is acknowledged.

All images © European Union 2016

Contents

Contents	2
Acknowledgements	4
Abstract	5
1. Introduction	6
2. Requirements for standardisation	6
2.1. Case study in post-event video analysis	7
2.1.1. Interoperability aspects	9
2.1.1.1. Level 1 — Technical interoperability	9
2.1.1.2. Level 2 — Syntactic interoperability	10
2.1.1.3. Level 3 — Semantic interoperability	10
2.1.1.4. Level 4 — Pragmatic interoperability	11
3. State of the art on standardisation activities and roadmap	11
3.1. Existing standardisation activities	11
3.1.1. Interface standards organisations	11
3.1.1.1. ONVIF	11
3.1.1.2. PSIA	11
3.1.2. TCs and WGs	12
3.1.2.1. ISO/TC 292	12
3.1.2.2. ISO/TC 223	12
3.1.2.3. CEW/TC 391	12
3.1.2.4. ANFOR Group	12
3.1.2.5. CLC/TC 79 WG12	13
3.1.2.6. EC TC 9 WG46	13
3.1.2.7. PSIA WG on video analytics	13
3.1.2.8. Video and image analytics coordination group	14
3.1.3. Relevant ISO/IEC standards	14
3.1.3.1. Alarm systems	15
3.1.3.1.1. Cenelec — EN 50132-1:2010	15
3.1.3.2. Multimedia	15
3.1.3.2.1. ISO/IEC 23000-10:2012	15
3.1.3.3. Forensics	16
3.1.3.3.1. ISO/IEC 27037:2012, 27041:2015, 27042:2015, 27043:2015, 27050 (draft)	16
3.1.3.4. Video surveillance	16
3.1.3.4.1. ISO 22311:2013	16
3.1.3.4.2. NF EN 62676:2014	17
3.1.3.5. Risk analysis	17

3.1.3.5.1. ISO 31000:2009 — Risk management — Principles and guidelines	17
3.1.4. Benchmarking activities	18
3.1.4.1. Initiatives	19
3.1.4.1.1. iLids (UK)	19
3.1.4.1.2. Etiseo (FR)	19
3.1.4.1.3. PETS (UK)	19
3.1.4.1.4. VOTS (SI)	20
3.1.4.1.5. MOT (CH)	20
3.1.4.1.6. DARPA: Mind's Eye (US)	20
3.1.4.1.7. TRECVid (US)	20
3.1.4.1.8. MCSPT (US)	21
3.1.4.1.9. Visitors (US)	21
3.1.4.1.10. VACE (US)	21
3.1.4.1.11. CLEAR (US)	22
3.1.4.1.12. VERAAE (US)	22
3.1.4.1.13. VTAPS (US)	22
3.1.4.1.14. VAPS (US)	23
3.1.4.1.15. IARPA DIVA (US)	23
3.1.4.2. Projects	24
3.1.4.2.1. Subito	24
3.1.4.2.2. Protectrail	24
3.1.4.2.3. SECUR-ED	24
3.1.4.2.4. ARENA	25
3.1.4.2.5. Savasa	25
3.1.4.2.6. Advise	25
3.1.4.2.7. Forensor	25
3.1.4.2.8. ITEA 2 LINDO	25
3.1.4.2.9. Reveal	26
3.1.4.2.10. Caretaker	26
3.1.4.2.11. CRISP	26
3.1.4.3. Representative benchmark datasets	26
3.1.4.3.1. Caviar	27
3.1.4.3.2. Etiseo	27
3.1.4.3.3. PETS	27
3.1.4.3.4. iLids	27
3.1.4.3.5. VIRAT	28
3.1.4.4. Metrics	28
3.1.4.5. Ground truth	28

3.1.5. Certification of surveillance systems	28
3.2. Gap analysis	28
3.2.1. Interoperability	28
3.2.1. Certification	29
3.2.2. Benchmarking methodology	30
3.2.3. EU projects	29
3.2.3.1. Protectrail	29
3.2.3.2. Savasa	31
3.2.3.3. Advise	31
4. Basic elements for standardisation process concerning use of video surveillance systems	31
4.1. Standardisation process	31
4.2. Recommendations for video surveillance standards	32
5. Roadmap to achieve draft standard agreement concerning new standards in, and certification of, video surveillance systems	33
5.1. New standards in surveillance of critical infrastructure	33
5.2. Certification of surveillance systems for protection of critical infrastructure	33
6. Conclusions	34
6.1. Next steps for TG-VS	34
References	35
List of abbreviations and definitions	38
List of figures	39

Acknowledgements

The author gratefully acknowledges the contributions and reviews of the other members of the European Reference Network for Critical Infrastructure Protection (ERNCIP) Thematic Group Video Analytics and Surveillance, and the ERNCIP office.

Abstract

This report has been generated by the ERNCIP Thematic Group on Video Surveillance for Security of Critical Infrastructure (TG-VS).

It is widely recognised that standards play a major role as fundamental building blocks in product development, to ensure uniform quality in provision of services, and wider still in enabling the European Union (EU) security industry to be more competitive globally. However, there exist very few standards in the security domain and, in particular, in video surveillance systems.

The purpose of this document is to provide an overview of standards in video surveillance, including the need for standards, an overview of existing relevant standardisation efforts including gaps, and a roadmap for future standards development.

The first part of the document identifies the need for standardisation. The programming mandate M/487 issued by the EC to standards bodies in 2011 to study existing standards in the security domain, to establish gaps, and to propose a standardisation work programme, did not directly address video analytics or video surveillance in its remit, hence these areas remain weakly addressed. The report provides a case study on post-event investigative video analysis illustrating the requirements for such standards, especially on interoperability aspects.

The report then details standard development organisations relevant to surveillance, including an overview of their work. This includes interface standards organisations, working groups (WGs) and technical committees (TCs), and relevant ISO/IEC standards including on alarm systems, multimedia, forensics, video surveillance, and risk analysis. Standardisation also covers benchmarking activities, whose aim is to effectively assess the performance of algorithms and systems in order to attain robustness under varying conditions. The history of relevant initiatives is provided in this report, including details on specific programmes, projects, and methodology. Prior work on certification on video surveillance systems is also described.

The report then details a gap analysis of standards in video surveillance. This includes lack of standards at different levels of interoperability, lack of a universally agreed set of performance evaluation benchmarking metrics for video analytics, and the lack of European level certification for surveillance systems or its components.

Finally, the report makes a number of recommendations for video surveillance standards. This includes new work items to (1) develop one or more EU standards for surveillance of critical infrastructure, and (2) to develop a harmonised certification procedure for video surveillance systems and components for protection of critical infrastructure at EU level. Specifically, the following concrete actions by TG-VS would further standards development in video analytics and surveillance for critical infrastructure protection:

- develop a procurement framework to be used by critical infrastructure end users when procuring video analytics;
- undertake preparatory work in the development of one or more new standards in video surveillance;
- undertake preparatory (pre-normative) work in the certification of surveillance systems.

The audience for this report include EU level policy authorities (especially the Directorate-General for Migration and Home Affairs), industry, end users and other stakeholders with interest in deployment of video analytic and surveillance systems and methods, and academics researching surveillance and standardisation.

1. Introduction

Video surveillance plays a major role in protection of critical infrastructure. Whether for preventive or post-event analysis, authorities require the capability to rapidly exploit closed-circuit television (CCTV) data collected in the vicinity of infrastructure to understand situations and scenarios as they unfold, which are directly related to protection of the infrastructure.

It is widely recognised that standards play a major role in enabling interoperability, uniform quality in provision of services, reduction in costs, future-proofing, and wider still, in enabling the EU security industry to be more competitive globally.

However, there exist very few standards in the security domain and, in particular, in video surveillance systems. Specifically, there is a lack of specific standards associated with the description of archived video content for video-based security systems. This becomes an issue when retrieving information from disparate systems, as there is no common language that these types of security systems describe their information in. This problem is enlarged by the large number of current vendors of video based detection systems (VBDS), and the large installed base of legacy systems.

The purpose of this document is to provide an introduction to standards in video surveillance, including the need for standards, an overview of existing relevant standardisation efforts including gaps, and a roadmap for future standards development.

In particular, the report provides a gap analysis of standards in video surveillance. This includes details on the current lack of standards at different levels of interoperability, lack of a universally agreed set of performance evaluation benchmarking metrics for video analytics, and the lack of European level certification for surveillance systems or its components.

To address these issues, the report makes a number of recommendations for video surveillance standards. This includes new work items to (1) develop one or more EU standards for surveillance of critical infrastructure, and (2) to develop a harmonised certification procedure for video surveillance systems and components for protection of critical infrastructure at EU level.

The audience for this report include EU level policy authorities (especially the Directorate-General for Migration and Home Affairs), industry, end users and other stakeholders with interest in deployment of video analytic and surveillance systems and methods, and academics researching surveillance and standardisation.

2. Requirements for standardisation

Standards development includes establishment of consistent protocols that are universally understood and adopted. Standards are vital for interoperability of technologies used by law enforcement and other authorities. However, to date, very few EU wide standards exist in the area of security, especially in surveillance and video analytics. This is largely due to lack of agreement on processes and best practice in the heterogeneous security market, which has resulted in divergent national standards.

In 2011, underpinned by a number of studies including the European Society Research and Innovation Forum, the European Commission announced in its Communication on a Strategic Vision for European Standards the need to speed up standardisation efforts in the civil security area (Poustourli and Kourti 2014). A programming mandate M/487 (European Commission 2011) was subsequently issued to the European Standardization Organizations (CEN, Cenelec and the European Telecommunications Standards Institute (ETSI)) to obtain a detailed overview of existing international, European and national standards in the security area, as well as to set out a list of standardisation gaps and to propose a standardisation work programme.

The work was accepted by the European Standards Organizations and allocated to CEN/TC 391 'Societal and Citizen Security'. As a result, a study was carried out to analyse the current security standardisation landscape and the security end-users needs of standards in three thematic areas: chemical, biological, radiological, nuclear and explosives, border security — automated border control systems, as well as biometric identifiers; and crisis management and civil protection — including communication and organisational interoperability. The outcome of the study was a set of roadmaps in each of these areas and ultimately a set of priorities in the work programme of the CEN and Cenelec TCs.

The CEN/TC 391 study did not consider video surveillance or video analytics and hence these areas remain weakly addressed from a standardisation roadmap perspective. In particular while low-level general standards (for example, the Moving Picture Expert Group (MPEG)) (Gao, et al. 2013) (Zhang, et al. 2013) (Institute of Electrical and Electronics Engineers (IEEE) Computer Society 2014) on video coding are well established for universal use, agreement on standardisation in video surveillance formats used by video surveillance systems is lacking. Moreover, the problem becomes even more severe when considering, for example, metadata (J. van Rest 2011) (van Rest, et al. 2014) generated by video analytic methods such as tracking and event/behavioural analysis.

Apart from interoperability aspects related to both real-time and post-event analysis (Leal-Taixé, et al. 2015), there is also a need to benchmark performance of video surveillance systems. Performance evaluation refers to the process of assessing quantitatively (as well as qualitatively) performance of individual video analytic methods (for example, detection, tracking, recognition, event or behavioural analysis) against the ground truth, using a set of appropriately defined metrics. To date, there is no universally agreed set of metrics or overall benchmarking process. Moreover, while initiatives such as the Imagery Library for Intelligent Detection Systems (iLids) (Sage, Nilski and Sillett 2010) have also addressed the related area of certification of video analytics systems for specific scenarios, the approach has not been universally adopted for a wider range of video analytic methods and/or scenarios, nor at a European level.

Finally, in addition to technical interoperability and benchmarking aspects, there is also a demand for transparency in relation to legal and ethical aspects of surveillance (Savasa 2014). For example, 'privacy-by-design should facilitate cross country acceptance of surveillance products. But how?' (Gemo and Andritsos 2011).

It should be noted that this report does not focus on legal and ethical issues. Emphasis is paid to interoperability and benchmarking aspects of surveillance systems and methods.

2.1. Case study in post-event video analysis

This section provides a case study on post-event (forensic) video analytics (otherwise known as *a posteriori* analysis or video archive search) highlighting the process, issues and challenges around effective exploitation of video data for investigative purposes.

Imagine that a criminal or terrorism act has been committed in or around a critical infrastructure such as a power plant, air traffic control or transportation hub. Most likely, surveillance cameras will have recorded the act and the circumstances before and after, as well as the surrounding areas. Tasks including tracking a person through this network of CCTV cameras, or detection of an airborne drone, ought to be possible, and relatively easy, with the modern automated video analytics tools that are available today.

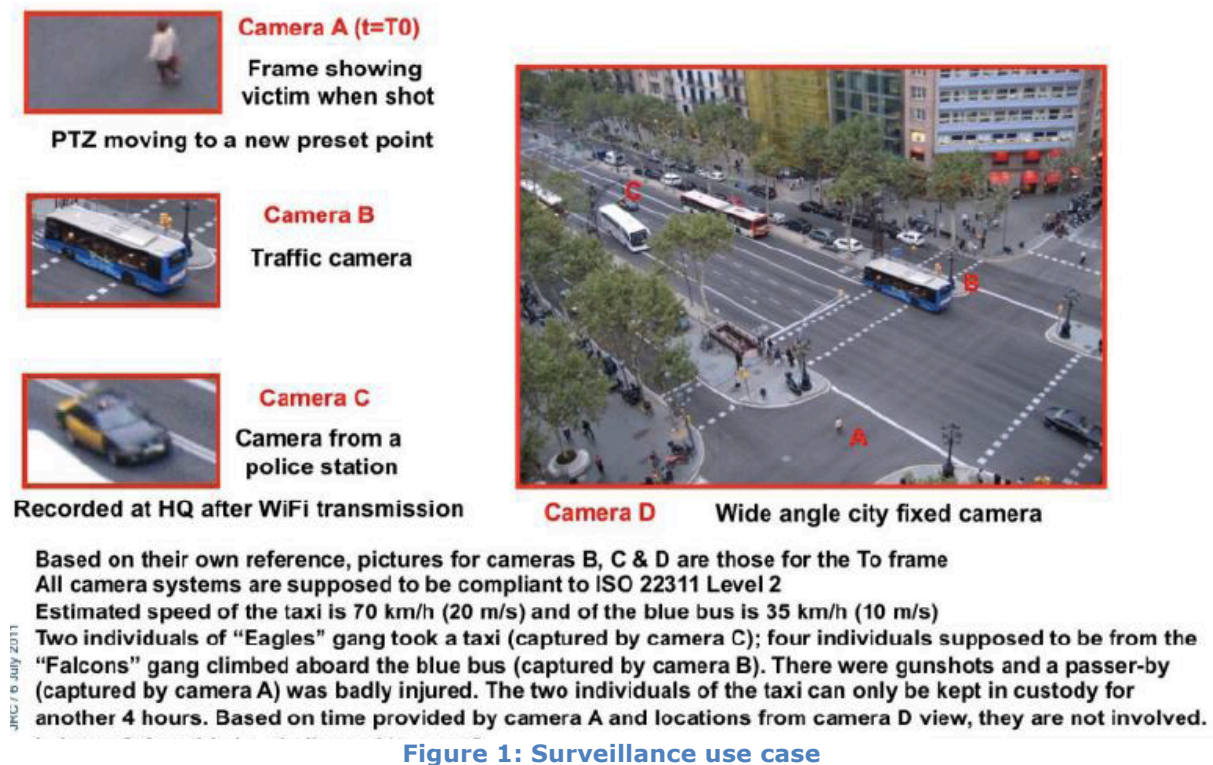


Figure 1 (Gemo and Andritsos 2011) shows an example of such an act. Based on analysis of the events conclusions can be reached. However, is it a true statement of events? Could it be challenged in court?

In reality, law enforcement agencies and the criminal justice system rarely use such tools, even though CCTV evidence is a valuable and compelling tool in the resolution of many cases. The analysis of large volumes of CCTV is a difficult and time-consuming process. This is because of interoperability problems, on multiple levels, between the various surveillance systems on the market. The video material from the site of, for example, a train station, where the act was committed is likely stored in a proprietary file format, using a semi-proprietary compression format, readable only using a proprietary software, and exportable to still images in a heavily compressed format. In short, there is a multitude of digital recording formats with poor interoperability and accessibility. Law enforcement face many interoperability problems in gaining effective access to the recordings and then it is a manual process of directly viewing the video to extract information. As a result, the overwhelming proportion of recorded CCTV is simply over-written with no information ever being extracted from it, thus undermining the value and potential of the investment in video sensors, storage and infrastructure.

In those cases where imagery is readily available (i.e., surveillance of large infrastructures), interoperability is lacking on higher levels. In order to really exploit the availability of large amounts of video data, it is necessary to have efficient and effective methods for searching the data. This requires that the video data should be tagged with metadata that describes it. There are several different kinds of metadata. The simplest are date of creation, camera type used and similar. In addition to such, metadata can be created using video analytics of varying complexity, ranging from simple features (e.g. motion present in the scene) to metadata generated by advanced computer vision algorithms (e.g. type of events and tracks (IEEE Computer Society 2014) and behaviour of persons). However, metadata is currently system specific. In summary, metadata, database structures, image quality (or, rather, the lack of these) are thus factors that can prevent the efficient use of available surveillance data.

In order to effectively exploit the large volumes of CCTV imagery it is necessary to develop and adopt a range of standards (Dufour 2012) (British Standards Institute (BSI) 2011). Currently, there exist national (UKGOV 2015), European and international standards for installers, maintainers and manufacturers of CCTV. These standards are limited, however, in while they largely focus on CCTV installation, transmission and monitoring of CCTV for human monitoring, and includes some recommendations for effective image export capability and an easily replayable data format, they do not address automated surveillance and video analytics.

2.1.1. Interoperability aspects

The degree to which systems are interoperable can be quantified by using the interoperability levels depicted by the levels of conceptual interoperability model, see Figure 2 (Tolk, Diallo and Turnitsa 2007). The current issues in the domain of video surveillance are concentrated in the subset 1 to 4 of these levels.

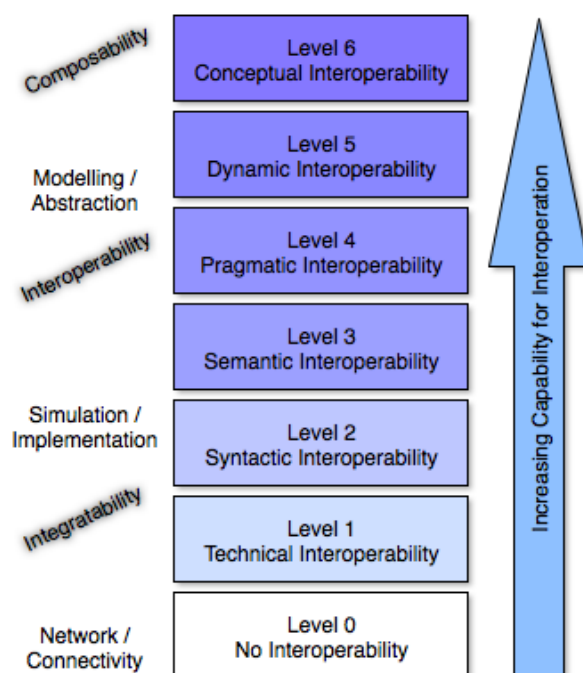


Figure 2: Levels of interoperability

2.1.1.1. Level 1 – Technical interoperability

At Level 1 the physical communication medium between video surveillance systems is considered. This is for example the cable, connector and storage medium, including the system clocks. Many different storage and transmission solutions are used. In the United Kingdom (UK), for example, the move towards IP-based system with networked cameras has been talked about for some years, but the UK police rarely encounters recordings from such systems — they are almost never present on the premises from which CCTV is retrieved. Other European countries are 'catching up' in the amount of CCTV systems, and a side effect of this is that for example in the Netherlands and Sweden the ratio of IP-based systems versus analogue systems is higher than in the UK. The purchasing decision for a CCTV system is mainly based on price level, but may be influenced by industry guidelines (such that the British Security Industry Association (UK) or HetCCV (the Netherlands) have developed) or existing norms. A preliminary study of these guides shows that they neglect the demands of the forensics use case. For example, they do not mention the need of physical connectors to be accessible by the police, let alone the type and format of the data going through such connectors. From a forensics

point of view, there is also a risk in specifying such a data format, because that might imply lossy transcoding of an internal format to that external format. These guidelines are not mandatory. A law, or a demand, from the insurance company might force the adoption of such best practices. However, there already is a huge installed base. Because of this, the police have strict procedures for recovering data from existing CCTV systems. When data from a CCTV system is retrieved, the time settings of recorders are referenced to the speaking clock so that any offset in date and time is known. For serious crimes where CCTV is relevant, the police may take whole CCTV recorders, but usually they try to be as specific as possible. The UK Metropolitan Police Services Video Laboratory has a 'factory process' for extracting the CCTV data so that the investigation teams have fast access to it.

2.1.1.2. Level 2 — Syntactic interoperability

Level 2 concerns video formats and protocols. These formats and protocols are tightly linked to the technical interoperability of Level 1. Proprietary data formats and even closed data formats are very common. Even within a format, many configurations can be used. In a typical CCTV installation, these are typically determined by other use cases than the forensics use case. For example, live viewing of CCTV is done for crime prevention. This means that data formats that optimise fluidity over resolution are more likely to be used in such installations. Police video laboratories deal with a range of CCTV video data, with differing resolution for example, PAL (768 x 576) or HD (1280 x 730) pixels resolution. It is very tempting to make one data format mandatory for all CCTV systems. This is not a viable strategy for the following reasons: (i) it will hamper technological progress; (ii) it neglects other use cases for CCTV surveillance; (iii) it neglects the vast installed base and (iv) this decision is outside of the scope of any country. The other option is to produce a universal data format decoder.

2.1.1.3. Level 3 — Semantic interoperability

Level 3 is concerned with the way the video is described by video analytics and by manual annotation. Two systems are semantically interoperable if the intended meaning of the data and information created in one of them can be successfully transferred to the other. The state of the art in data formats can be found in several ontologies that are already available for partial descriptions, such as the Open Network Video Interface Forum (ONVIF) 1.0, MPEG7, and the ontologies of Performance Evaluation of Tracking and Surveillance (PETS), the TREC Video Retrieval Evaluation (TRECVID) and iLids (Sage, Nilski and Sillett 2010). There is no single data format that is useful for all domains where CCTV is used and hence there has been a recommendation to create one (J. van Rest, Surveillance and video analytics: factors influencing the performance 2015). Furthermore, van Rest, et al. (2014) concluded in their study on metadata that MPEG7 and the Surveillance Application Metadata (SAM) (Schallauer, et al. 2009) were the best candidates to build such a standard on. It can be postulated that ONVIF will create this standard with one of the next versions, however again one has the problem of the huge installed base. One can distinguish two main branches of methods to describe the content of video: manually (by human eyeballing) or automatically by using video analytics. Because both have their drawbacks, a combination usually works best, but which combination that is, depends heavily on the actual case and available resources: Are trained operators available? Are the video analytics algorithms suitable for this camera position and orientation? When the annotation is performed by automated systems, known as video analytics or video content analysis, the output data format is usually determined by the choice of video analytics vendor. For example, the Milestone Alert Data format is yet another format — even though an open format — that is different from the previous mentioned formats. Most products in this area are integrated with video viewers or advanced alarming systems and the semantic data formats are often hidden internally.

2.1.1.4. Level 4 — Pragmatic interoperability

Level 4 is concerned with the way algorithms, methods and procedures are applied. Choosing between several types of video analytics (including a human) to perform a certain search query depends on knowledge about such methods in relation to the characteristics of the video, the sensor and the context of the recording. For example, the TRECVID, PETS and iLids efforts (see below) describe the expected performance of such systems on specific scenarios. Being able to select the most relevant algorithm and user interface for a specific combination of query and type of video, sensor and context is the goal on this level. The data format is one consideration, the way it is populated with data quite another. In the forensics use case, one is dependent on previously made decisions on the lower levels of interoperability: camera selection, camera position, lighting conditions, video frame rate, compression ratio, etc. The fact that one has a video stream is the only assumption one may make at this point. In general, automatic processing of CCTV recordings in the forensics use case is very hard. As well as time lapse, small targets, and poor quality images, there are often lighting changes and panning cameras. The type of issues that are important may also be very specific to an investigation.

3. State of the art on standardisation activities and roadmap

3.1. Existing standardisation activities

The following sections detail standard development organisations relevant to surveillance, including an overview of their work.

3.1.1. Interface standards organisations

For IP-based physical security surveillance products two organisations have been working towards the creation of interoperability specifications, ONVIF and the Physical Security Interoperability Alliance (PSIA). Neither organisation has yet dominated the arena however it is speculated that they will merge in the future. Currently many vendors support interoperability specifications from both ONVIF and PSIA in their products.

3.1.1.1. ONVIF

ONVIF is an established open industry forum and non-profit organisation, founded in 2008 by Axis, Bosch and Sony (currently over 500 member organisations), for the development of a global standard for the interface of IP-based physical security products. The aim of ONVIF is to ensure interoperability between products regardless of manufacturer by creating an open standard for communication between IP-based physical security devices (e.g., surveillance cameras). The ONVIF specification defines a common protocol for the exchange of information between network video devices including automatic device discovery, video streaming and intelligence metadata. The focus of ONVIF is on real-time streaming of data and metadata. As of February 2015, ONVIF has three specialist profiles (Profile S for streaming video; Profile G for recording and storage; Profile C for physical access control) and Release Candidates Profile A, for access control configuration, and Profile Q, for easy installation and advanced security features), but none dedicated to video analytics.

3.1.1.2. PSIA

PSIA, founded in February 2008 and incorporated in March 2009, is represented by a global consortium of over 65 physical security manufacturers and integrators including Honeywell, Verint, IBM, Cisco and GE, focused on promoting interoperability of IP-enabled devices. Compared to ONVIF, PSIA have taken a more holistic and systems-based approach to standardisation, with five active WGs including one on video analytics, detailed below. In 2010, PSIA released a Recording and Content Management Specification describing standards for recording, managing, searching, describing, and streaming multimedia information over IP networks. The specification includes XML

Schema Definitions and XML examples to aid development of standards-based products. In 2010, PSIA released a Video Analytics Specification, which specifies an interface that enables IP devices and video management/surveillance systems to communicate video analytics data in a standardised way. Most recently in February 2015, PSIA released an Area Control Specification, which specifies the communication into access control and intrusion products, making them interoperable with the overall security system.

3.1.2. TCs and WGs

3.1.2.1. ISO/TC 292

The international ISO/TC 292 TC on 'Security and resilience' was established on 1 January 2015 and involves over 50 countries. It works with standardisation in the field of security to enhance the safety and resilience of society. The committee is responsible for more than 20 published international standards. Six WGs have been set up to conduct the work. ISO/TC 292 is responsible for wide range of standards and other documents including on 'business continuity management', 'emergency management', 'community resilience', 'authenticity, integrity and trust for products and documents', and 'protective security'. With respect to the ERNCIP TG-VS the most relevant standard under development is 'Protective security — ISO 22311 — Societal security: Video surveillance — Export interoperability' detailed below.

3.1.2.2. ISO/TC 223

ISO/TC 223 was set up in 2001 to develop international standards that aim to increase societal security, specifically the protection of society from and response to incidents, emergencies, and disasters caused by intentional and unintentional human acts, natural hazards, and technical failures. Since 2001, ISO/TC 223 has developed a range of standards in societal security and is strongly linked to ISO/TC 292. Specifically, as a result of WG5 (of 6) devoted to public safety (led by Jean-Francois Sulzer, Thales), ISO/TC 223 has published as part of its work ISO 22311 — Societal security: Video surveillance — Export interoperability (see below). Note that WG5 has been dissolved since completing its work. In addition to the standards, a series of technical specifications, reports and publicly available specifications have also been published.

3.1.2.3. CEN/TC 391

The main objective of European TC CEN/TC 391 (Societal and Citizen Security) is to elaborate a family of European standards, standard-like documents (e.g. procedures, guidelines, best practices, minimal codes of practice and similar recommendations) in the Societal and Citizen Security sector including aspects of prevention, response, mitigation, continuity and recovery before, during and after a destabilising or disruptive event. In particular, it was planned that a minimum format standard for security events will be developed within the scope of EC mandate M487 of CEN/TC 391.

3.1.2.4. ANFOR Group

ANFOR is an international group composed of association and subsidiaries and with the general aim to serve general interest and economic development of organisations. The Security forum of ANFOR — French branch of ISO/TC 223 dedicated to public safety — brings together all actors in field of security including infrastructure and state services. In November 2008, ANFOR made a proposal to ISO to set up an international WG with the aim to define minimum conditions for interoperability needed to exploit videos from different sources as directly as possible. Such a shared understanding of video content would rely on existing efforts/norms, including:

- video in general moving picture experts group: MPEG4, H264 of ISO IEC JTC 1 SC29
- digital TV: society of motion picture and television engineers (SMPTE)

- North Atlantic Treaty Organization (NATO): interoperability mechanisms for animated images STANdardisation Agreement (Stanag 4609)

The WG aimed to address (Dufour 2012):

- format for video content compression with quality required for exploitation in forensic police work, with preferred profiles, based on MPEG4 H264 from ISO CEI JTC SC29;
- minimum list of data describing the conditions of capture (i.e. metadata) for recording time and date of sequence capture (as well as camera field of view, zoom, Global Positioning System (GPS) coordinates,...);
- synchronisation of various elements captured at the same time (video, audio, metadata, alarms), with a recommended solution of ISO/CEI23000-10;
- format or transfer protocol enabling person exploiting videos to be aware of what form the content will be sent to them;
- integration of constraints relating to security and authentication of content that is evidential in court of law.

The project resulted in the standard ISO 22311 being published in October 2013 (see below).

3.1.2.5. CLC/TC 79 WG12

The scope of CLC/TC 79 — Alarm and electronic security systems — is to prepare international standards for the protection of buildings, persons, areas and properties against fraudulent actions having the purpose to enter in a place or to take or to use something without permission and other threat related to persons. The scope includes, under WG12, video surveillance systems (formally CCTV under WG7, and in addition to access control systems, fire detection and fire alarm systems, etc.) for security applications. Specifically, EC 62676-5 addresses video surveillance systems for use in security applications — Part 5: Data specifications and image quality performance for camera devices (see below for further details.) In August 2016, a new work item has been specifically proposed on video analytics under WG12.

3.1.2.6. EC TC 9 WG46

The scope of EC TC 9 — Electrical equipment and systems for railways — is to prepare international standards for the railways field which includes rolling stock, fixed installations, management systems (including communication, signalling and processing systems) for railway operation, their interfaces and their ecological environment. The relevance to the ERNCIP TG-VS is that WG46 addresses 'Onboard multimedia systems for railways' with the scope to define and implement a multimedia framework that includes the standardisation of a number of subsystems within the train that communicate using the subsystem defined by WG43 (Fadin and Umiliacchi 2011). One of the subsystems considered is 'Video surveillance/CCTV'. Overall, the work contributes to the development of the IEC 62580 standards, which includes for video surveillance/CCTV, IEC 62580-2 (Electronic railway equipment — On-board multimedia and telematics subsystems for railway — Part 2: Video surveillance/CCTV services) (edition 1.0) that is forecast to be published in July 2016.

3.1.2.7. PSIA WG on video analytics

In September 2010, the PSIA WG on video analytics released the Video Analytics 1.0 Specification based on ObjectVideo's OV Ready protocol. The main features of the open specification are:

- enables video analytics to more easily and consistently integrate with video management systems and physical security software platforms through standard interfaces;

- defines a standard way to share video analytics capabilities supported by an intelligent device and output, receive, store and use various video analytic events;
- open interface addresses event output including security alerts, counting events and analytics system health messages. The interface also supports the streaming of object metadata output, which includes foundational analytic output regarding all objects tracked by the analytics, including object classification, bounding box data and velocities.

The scope for the initial release of the Specification focuses entirely on video analytics capabilities discovery and analytic data output. Video analytic capabilities discovery will include standard configuration data exchange to enable any analytic device to communicate to another device or application its basic analytic capabilities at the device level and the video channel level (for multi-channel devices). This includes information such as the PSIA VAS version number supported, analytic vendor information (name, software version number, etc.), event types and mechanisms supported, and other supported configurations. From an analytic output perspective, the v1.0 Specification includes the definition of multiple types of analytic events, including alerts and counts, as well as video analytics metadata output.

3.1.2.8. Video and image analytics coordination group

The VIA-CG, part of the Networking and Information Technology Research and Development (NITRD) Program (US), was created to ensure and maximise successful coordination and collaboration across the U.S. Federal government in the important growing area of video and image analytics. It is chaired by John Garofolo at NIST and adopted the overall mission to 'provide a vehicle for visible world video and image analysis technology R & D strategy development, collaboration, and resource-sharing across the Federal Government and engagement mechanism with state, local, and tribal governments in applications spanning a variety of domains of national importance'. VIA-CG has six goals one of which is to identify standards, measurement, and R & D resources and needs.

3.1.3. Relevant ISO/IEC standards

The following sections outline the main standards, which relate to video analytics and critical infrastructure protection. The main categories of standards considered are: (1) alarm systems, (2) multimedia, (3) forensics, (4) video surveillance, and (5) risk management.

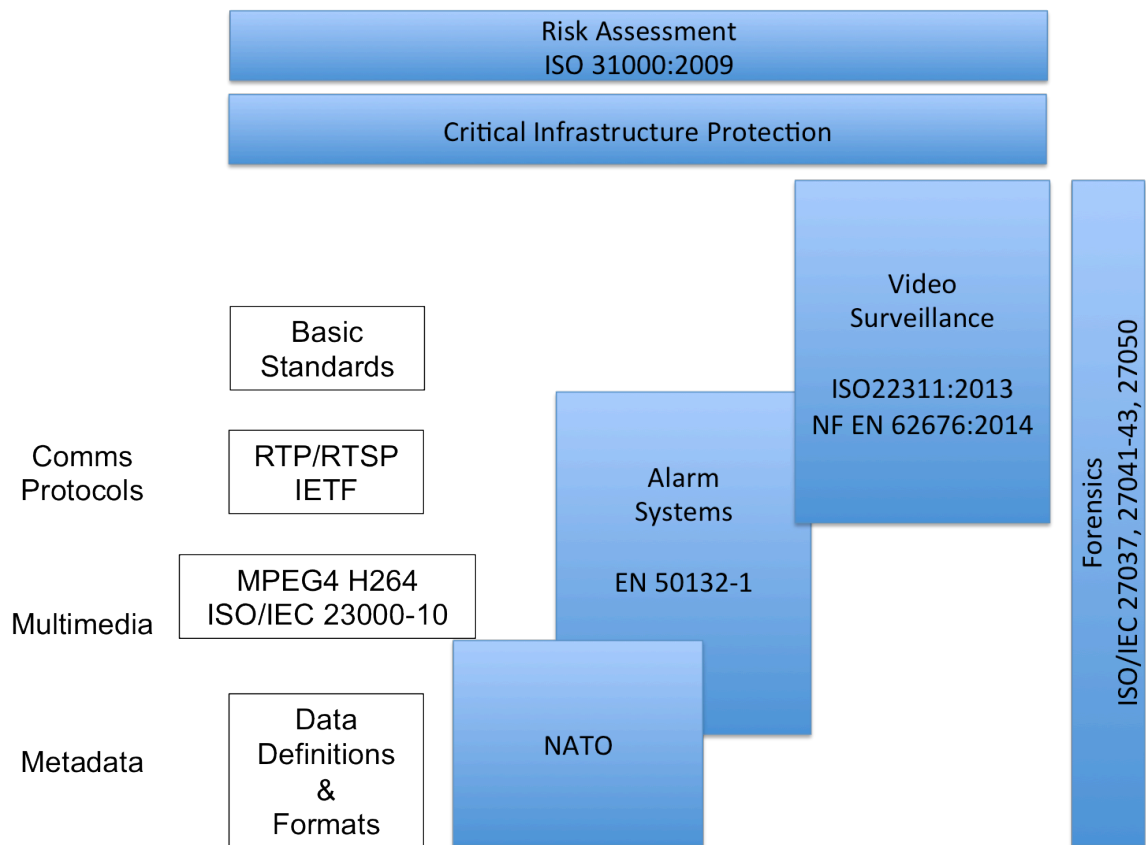


Figure 3: Main standards relating to video surveillance for protection of critical infrastructure (adapted from Sulzer (2014))

3.1.3.1. Alarm systems

3.1.3.1.1. Cenelec — EN 50132-1:2010

The European Norm 50132-1 'Alarm systems — CCTV surveillance systems for use in security applications — Part 1: System requirements' was established with the aim of ensuring a high consistent level of performance of video surveillance systems in Europe. The latter parts gave recommendations for the selection, planning and installation of CCTV systems comprising of camera(s) with monitor(s) and/or video recorder(s), switching, control and ancillary equipment for use in security applications (part 7) and video transmission (part 5). Note that 50132-1 has now been withdrawn and superseded by EN 62676-1-1 detailed below.

3.1.3.2. Multimedia

3.1.3.2.1. ISO/IEC 23000-10:2012

ISO/IEC 23000-10 (Information technology — Multimedia application format (MPEG-A) — Part 10: Surveillance application format) is an international ISO standard published in December 2012. ISO/IEC 23000-10:2012 specifies 'a file format designed to store data in and exchange data between surveillance systems. The file format provides an overall structure for media content and associated metadata. Media data coverage includes image, video and audio data. Specific features to support application of the format in surveillance systems include dedicated time information in a separate track as well as segmentation and segment linking provisions for media data'.

3.1.3.3. Forensics

3.1.3.3.1. ISO/IEC 27037:2012, 27041:2015, 27042:2015, 27043:2015, 27050 (draft)

There are a number of relevant forensics standards (ISO 27037; 27041-27043; 27050) whose main purpose is to promote good practice in methods and processes for forensic capture and investigation of digital evidence. Specifically, it includes provision of guidance on digital still and video cameras (CCTV), amongst other devices and formats.

ISO 27037, published in October 2012, is focused on the initial *capture* and storage of the potential image and video evidence and not on the subsequent (automated) analysis of the evidence.

ISO 27041 (2015) provides guidance on general *assurance* aspects of digital forensics. Specifically, that forensic tools and methods are applied properly.

ISO 27042 (2015) covers the analysis and interpretation of digital evidence.

ISO 27050 (currently under draft) addresses *electronic discovery* within the collected forensic data. This includes the actual *processing* (analysis/search) of video data, which would include the application of video analytics.

3.1.3.4. Video surveillance

3.1.3.4.1. ISO 22311:2013

The ISO international standard on 'Societal security — Video surveillance Format — Export interoperability' led by Jean-Francois Sulzer (Thales) provides (ISO n.d.) (European Commission 2011) 'an export interoperability profile which constitutes the exchange format and minimum technical requirements that ensure that the digital video surveillance contents exported are compatible with the replay systems, establish an appropriate level of quality and contain all the context information (metadata) necessary for their processing'. The standard is motivated by the needs of law enforcement where 'the authorities [require the capability] to be able to rapidly use the data collected by different CCTV systems from given locations'. The aim was not to invent a new format, but to rely heavily on a blend of individual technical standards separately developed, concentrating on the minimum set of profiles required to achieve the objective: for video in general MPEG (MPEG-4 H264 of ISO JTC 1 SC29), for the world of digital television (range of norms SMPTE), the NATO's interoperability mechanisms for animated images Stanag 4609, etc. Furthermore, the standard would represent a nonproprietary affordable solution for all future systems. The standard includes the following video content components:

- format for video content compression with quality required for exploitation in forensic police work, with preferred profiles, based on MPEG-4 H264 from ISO CEI JTC SC29;
- minimum list of data describing the conditions of capture (i.e. metadata), the time and date the sequence was recorded, angles of view and the zoom value (for PTZ cameras), GPS coordinates for a camera on a vehicle, etc.; for each of these categories of metadata a specific means of representation (e.g. XML);
- means to precisely synchronise the various elements captured at the same time, such as video(s), sound, metadata and alarms; the recommended solution is the format MPEG-A (ISO/CEI 23000-10);
- format or transfer protocol enabling a person exploiting the videos to be aware of what form the content will be sent;
- means to integrate constraints relating to security and authentication of content that is valid as evidence in a court of law.

The standard was voted to Committee Draft status in November 2010 (ISO/CD 22311), approved by TC 223 in June 2012 and subsequently published by ISO in November 2012. The standard (as of April 2015) is designated to be revised.

3.1.3.4.2. NF EN 62676:2014

EN 62676-1-1 (Video surveillance systems for use in security applications, 2013) is a series of standards intended to enable flexibility to overcome problems a system designer may have. It should be noted that the BS EN 62676 series of standards are the first standards for CCTV video surveillance that will be used to any significant extent in Member States and include the use of security grading. The full set of standards is as follows:

- Part 1-1: System requirements — General specifies the minimum requirements and gives recommendations for video surveillance systems installed for security applications;
- Part 1-2: Video transmission — General video transmission — Requirements;
- Part 2-1: Video transmission protocols — General requirements;
- Part 2-2: Video transmission protocols — IP interoperability implementation based on HTTP and REST services;
- Part 2-3: Video transmission protocols — IP interoperability implementation based on Web services defines procedures for communication between network video clients and video transmitter devices based on Web services. This new set of specifications makes it possible to build network video systems with devices and receivers from different manufacturers using Web services. This international standard also contains full XML schema and Web Services Description Language definitions for the introduced network video services. Furthermore, appropriate protocol extensions have been introduced in order to make it possible for network video manufacturers to offer a fully standardised network video transfer solution to its customers and integrators;
- Part 3: Analog and digital video interfaces;
- Part 4: Application guidelines.

3.1.3.5. Risk analysis

3.1.3.5.1. ISO 31000:2009 — Risk management — Principles and guidelines

ISO 31000:2009 provides a high-level set of principles, framework and processes for managing risk and implementing risk management. The standard can be used by organisation irrespective of its size, activity or sector.

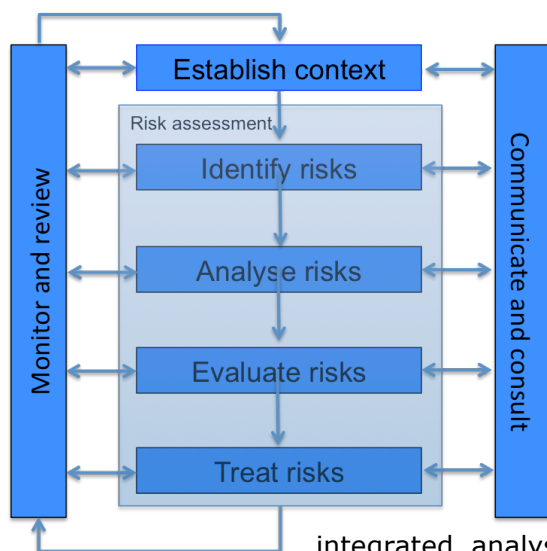


Figure 4: ISO 31000 risk management process

In terms of critical infrastructure, risks in relation to performance of visual surveillance methods and systems need to be considered as part of the overall risk management profile. In particular to consider standardisation of methods for visual surveillance as part of security management system certification. Saponi and Sciutto (2014) provide a detailed description of a method of risk analysis and assessment responding to the decision-making needs and which allows development of an

integrated analysis of the elements an organisation comprises (technological systems, procedures and human factor) through a multi-risk analysis aimed at assess technological failure, intentional attacks and natural disasters. This enables evaluation, for example, of the effects of the security surveillance systems on safety and vice versa (Saponi and Sciutto 2014).

Finally, as noted by Lewis (2014), 'ISO 31000 cannot be used for certification purposes, but some bodies do offer professional certification schemes for competence in implementing it' (Sapori and Sciutto 2014).

3.1.4. Benchmarking activities

It is not possible to consider standardisation in video surveillance and video analytics by considering interoperability aspects alone. There is a need to consider benchmarking activities whose aim is to effectively assess the performance of algorithms and systems in order to attain robustness under varying conditions (scene complexity, illumination, etc.).

A large number of algorithms have been designed and tested for the tasks of object detection and tracking as well as for detection of events of interest, abnormalities or criminal behaviours. However, despite this effort by the community, it is still difficult to compare or evaluate such algorithms because of the lack of standard metrics and benchmarks that indicate how detection, tracking and threat analysis system perform against a common database.

To answer the need of having a publicly available set of annotated video sequences, many evaluation programmes (including Etiseo, PETS, Caviar and TRECvid) have been created. Such research programmes provide video sequences at various difficulty levels together with associated ground truth. However, the same global difficulty levels may be constituted by different individual video processing problems (e.g. shadows, reflections, weak contrast, etc.). Consequently, the evaluation processes employed do not enable one to gain sufficient insight into each image processing algorithm. Specifically, for a given algorithm, the evaluation does not indicate which video processing problems that it has to pay attention to, which improvement is the most crucial and under which conditions this algorithm can achieve satisfactory performance, or is likely to fail to provide useful results. Standardisation has not been achieved to date either at the individual detection level or at the overall surveillance architecture level.

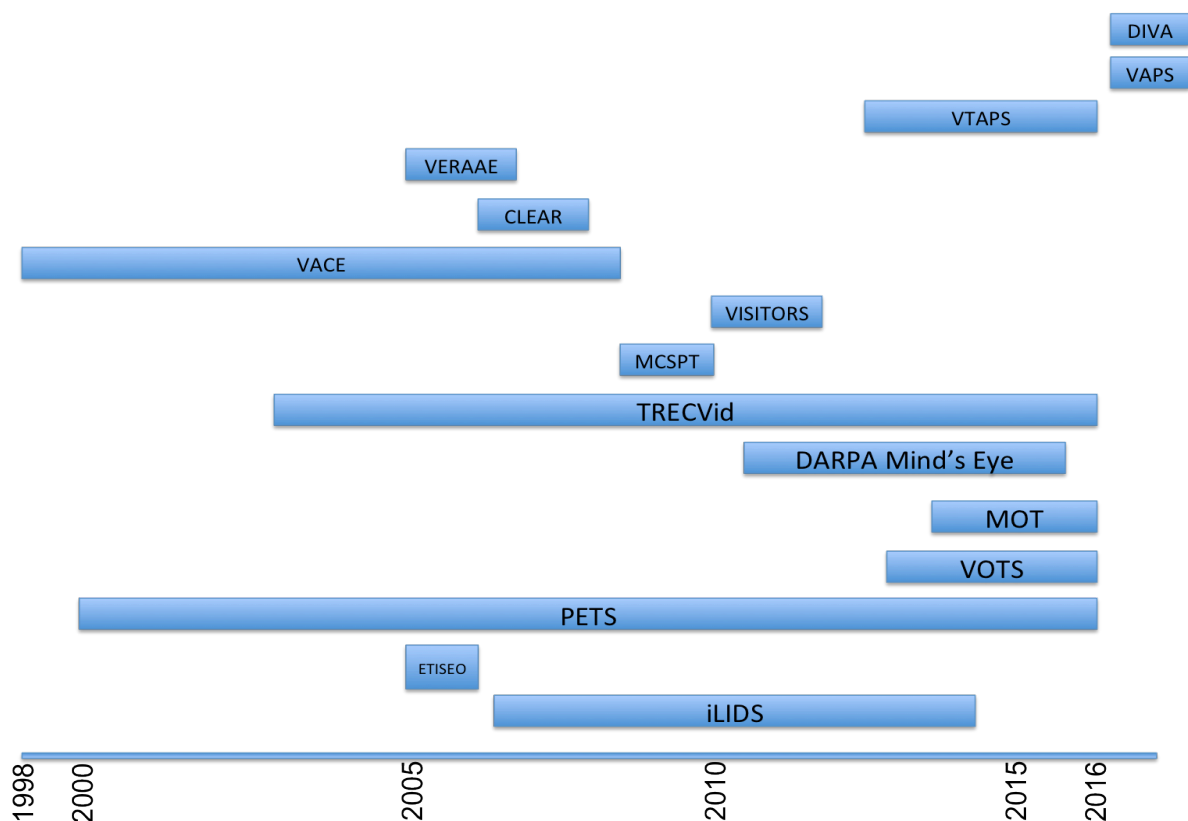


Figure 5: Timeline of evaluation initiatives

For a recent survey of datasets, performance metrics and approaches in video analytics evaluation, the reader is referred to Gorodnichy, Laganieri and Macrini (2014). Further, Marcenaro (2016), a publication of the ERNCIP TG-VS, details existing datasets for video analytics and determines how best to enable collection/common access to data sets in the EU for testing/evaluation of video surveillance software.

The following sections summarise the main benchmarking activities (datasets, metrics, evaluation protocols and challenges) that have been developed over the last 15 years.

3.1.4.1. Initiatives

3.1.4.1.1. iLids (UK)

iLids was launched as the UK government's (formerly Home Office Scientific Development's Branch, now Home Office Centre for Applied Science And Technology (CAST) benchmark for Video Analytic (VA) systems developed in partnership with the Centre for the Protection of National Infrastructure (CPNI) in November 2006. The benchmark consisted of a collection of real-world CCTV imagery with two aims: (1) performance evaluation of VBDS, and (2) imagery for future system and algorithm development. The benchmark consisted of public datasets for training and testing respectively and a private set for system evaluation by the Home Office. The programme implemented a certification approach for products, using in-house developed evaluation tools that met certain performance levels. CAST ceased distribution of the datasets in November 2014 with evaluation (certifications) continued to be offered on a commercial basis by BRE Global Ltd.

3.1.4.1.2. Etiseo (FR)

Etiseo (www-sop.inria.fr/stars/ETISEO/index.htm) is a Video Understanding Evaluation project which ran for 2 years between 2005-2006 and was part of the Techno-Vision evaluation network funded by the French Ministry of Defence and French Ministry of Research and coordinated by Silogic (FR). The aim was to evaluate vision techniques for video surveillance (focusing on pedestrian and vehicles), not so much from an end user point of view, but through study of the dependency between algorithms and the video characteristics. The Etiseo methodology addressed each video processing problem separately and proposed a set of evaluation metrics (as well as their benefits, limitations and conditions of use), automatic evaluation tools and visualisation tools. During the project 20 international teams submitted (e.g. tracking) results to the scientific leader (INRIA, FR) — i.e. a closed evaluation — which were then evaluated and presented at project meetings. Since the project ended, the evaluation data and tools have been made public. The main directions for future work were identified as: (1) consolidation of the evaluation activities, (2) standardisation, (3) organisation of several campaigns, (4) targeted use cases (e.g. metro) and specific themes (e.g. PTZ cameras, multi camera, and mobile cameras).

3.1.4.1.3. PETS (UK)

In 2000, based on the situation that the growth in the development of the video surveillance field has not been met with complementary systematic performance evaluation of developed techniques, the first PETS workshop was held in conjunction with FG'2000 and co-sponsored by the IEEE. The premise made was that it is especially difficult to make comparisons between algorithms if they have been tested on different datasets under widely varying conditions. The workshop was unique in that all participants were testing algorithms on the same dataset, published as part of the workshop. Results were submitted in XML format and evaluated by the workshop organiser, the University of Reading. Since 2000, PETS has become an established workshop series with new datasets collected and released largely on an annual basis. PETS is differentiated from other more recent evaluation efforts in that it is strongly scenario driven and usually linked to ongoing national or European project addressing tasks such as public space surveillance, critical infrastructure protection or maritime surveillance. Moreover, PETS has provided the wider community with a platform for

advancing performance evaluation methodology and standards, including development of datasets and metadata, metrics, and quantitative evaluation of detection and tracking capability. Since 2000, 18 PETS workshops have been held in conjunction with major computer vision conferences. The most recent workshop (PETS2016) was held in conjunction with CVPR2016.

3.1.4.1.4. VOTS (SI)

The Visual Object Tracking (VOT) (www.votchallenge.net) challenges provide the visual tracking community with a precisely defined and repeatable way of comparing short-term trackers as well as a common platform for discussing the evaluation and advancements made in the field of visual tracking. The focus is on single-target tracker performance evaluation using a cross-platform evaluation kit. Three benchmark challenges have been held to date — VOT2013, VOT2014 and VOT2015 with the latest dataset containing 60 short sequences and which attracted 62 trackers. The VOT2015 challenge also included a separate sub-challenge on thermal imagery (VOT-TIR2015) with 20 sequences. A multi-platform evaluation system allowing easy integration of third-party trackers has also been developed. The fourth benchmark challenge is scheduled to be held in October 2016 in Amsterdam in conjunction with the European Conference on Computer Vision (ECCV).

3.1.4.1.5. MOT (CH)

The Multiple Object Tracking (MOT) Benchmark (<https://motchallenge.net/>) (Leal-Taixé, et al. 2015) was set up in October 2014 to provide a unified framework for evaluation of multi-target people tracking. The benchmark provides:

- a large collection of datasets, some already in use and some new challenging sequences;
- detections for all the sequences;
- a common evaluation tool providing several measures, from recall to precision to running time;
- an easy way to compare the performance of state-of-the-art tracking methods;
- several challenges with subsets of data for specific tasks such as 3D tracking, surveillance, sports analysis.

MOT relies on *crowdsourcing*, and encourages researchers to submit their sequences to the benchmark, so the quality of MOT systems can keep increasing and tackling more challenging scenarios. A first workshop organised on the MOTChallenge benchmark took place in early 2015 in conjunction with the Winter Conference on Applications of Computer Vision, with only two entries received. A second workshop will be held in conjunction with the ECCV, Amsterdam, in October 2016.

3.1.4.1.6. DARPA: Mind's Eye (US)

The Defence Research Projects Agency (DARPA) Mind's Eye was instigated in September 2010 as a video analysis research programme focussed on advanced AI. In total 12 international research teams and three commercial integrators were involved in the 5-year programme. A number of collaborative projects were set up under the initiative and methods developed for video event recognition. Specifically, to include recognition of human activities in video and to predict what might happen next. Research also involved developing software to flag unusual events and deduce actions that may be occurring off-camera. In relation to standards developments, new datasets (www.visint.org) were produced and an ontology for human activities developed.

3.1.4.1.7. TRECvid (US)

The TRECvid series is sponsored by the National Institute of Standards and Technology (NIST) and other U.S. government agencies. Since 2003, it has promoted progress in content-based analysis of and retrieval from digital video via open, metrics-based evaluation. TRECvid is a laboratory-style evaluation that attempts to model real world situations or significant component tasks involved in such situations, including automatic

segmentation, indexing, summarisation and content-based retrieval of digital video broadcast news, documentary, and education programming. 2016 represents the 16th annual evaluation cycle and in which TRECVID will evaluate participating systems on six different video analysis and retrieval tasks using various types of real world datasets.

3.1.4.1.8. MCSPT (US)

The Multiple Camera Single Person Tracking Challenge Evaluation (MCSPT) was held for 2 years running (2009, 2010) in conjunction with the Advanced Video- and Signal-based Surveillance conference. Jointly sponsored by Home Office Scientific Development Branch, CPNI, and NIST, the goal was to facilitate research via a common evaluation task that focuses on one aspect of person tracking technologies: the ability to track a specified person within a video sensor field using a small set of *in situ* exemplar video images to specify the person. MCSPT referred to these technologies as Single Person Tracking technologies. The challenge was open to all interested participants and the results of the evaluations discussed during a special session at each of the respective events (Fadin and Umiliacchi 2011).

3.1.4.1.9. Visitors (US)

In 2010, NIST brought together major stakeholders from the retail and security industries, computer vision technologists/developers, the research community, law enforcement, and government agencies in a common mission to advance the state of the art in predictive video analytics. The focus of the Visitors project is to advance technologies and methodologies used to detect persons engaged in suspicious activities as applied in the retail domain. According to the project, Visitors goes a step beyond any video analytic work done to date, in the sense of promoting the advancement of video analytics that are able to predict that a criminal event is about to take place through the observation of specific behaviours. This is done by predictive analysis, using software to analyse a video stream and to identify suspicious behaviours that enable notification of security personnel while tracking the person of interest up to, and including, the shoplifting event. The Visitors project will require the development of a distributable video data corpus of normal and abnormal activities encountered in a retail environment. As of late 2010, three senior-level WGs were set up as follows.

- (1) Suspicious Behaviours, which encompasses law enforcement, psychologists, senior loss prevention officials, computer vision experts, and scientists. Intense brainstorming produced a list of salient external, internal and group behaviours which may be precursors to a shoplifting event.
- (2) Strategic Planning which is comprised of government only personnel and is looking at steps forward in evaluation protocols and planning.
- (3) Video Data Collection which met to discuss the various issues involved in collecting and disseminating video for training and testing algorithms. Options for using existing retail video databases and alternatively creating scenarios, based on desired suspicious behaviours, are being pursued. The group discusses a myriad number of issues, not least being dealing with legal issues concerning use of Human Subjects in research.

While the most recent activity in late 2010 included receiving clearances for data access and dissemination, development of scenario-based data collection protocols, and an evaluation plan and metrics for an initial pilot, there has been no evident progress on these tasks since.

3.1.4.1.10. VACE (US)

The Video Analysis and Content Extraction (VACE) was set up as one of three information exploitation programmes by the Advanced Research and Development Activity (ARDA) in Information Technology, created as a joint activity of the Intelligence Community and the Department of Defence in late November 1998. The goal was to develop novel algorithms for automatic video content extraction, multimodal fusion, and

event understanding. Funded in 3 2-year phases one of the domains considered was surveillance, specifically including people and vehicle detection and tracking, in which some progress was made. There was a focus on development of (VACE) diagnostic metrics, especially single-number metrics, which has remained as a significant legacy of the programme (Kasturi, et al. 2009).

3.1.4.1.11. CLEAR (US)

The Classification of Events, Activities, and Relationships (CLEAR) was a multi-national evaluation series, starting in 2006, that brought together researchers from the U.S. ARDA VACE programme (see above), the EU funded Computers in the Human Interactive Loop project, as well as the Augmented Multiparty Interaction programme, in a 2-year effort to:

- provide a common international evaluation framework for the perception of people, their activities, and interactions;
- to serve as a forum for the discussion and definition of related common benchmarks, including the definition of tasks, annotations, metrics and evaluation procedures.

The evaluation tasks included person tracking. Overall, the work resulted in evaluation packages for the various tasks, including datasets, annotations, scoring tools, evaluation protocols and metrics, which were made available through the Evaluations and Language Distribution Agency and NIST. In a similar outcome to VACE, the developed CLEAR evaluation metrics remain a significant legacy of the programme (Bernardin and Stiefelbogen 2008) (Kasturi, et al. 2009).

3.1.4.1.12. VERAAE (US)

The Video Event Recognition Algorithm Assessment Evaluation (VERAAE) was set up by US ARDA in 2005 as a comparative study of Video Event Recognition (VER) algorithms to assess the applicability, usefulness and limitations of different approaches. The motivation was that several promising VER approaches existed, but had varying degrees of success with different types of event detection. Further, there was no largely accepted criteria or dataset (with ground truth) and the performance of VER algorithms is highly dependent on the results of object detection and tracking, making evaluation of the event recognition alone very difficult.

3.1.4.1.13. VTAPS (US)

In July 2013, the U.S. Government identified a need for advanced video tracking analytic technology that works with the video generated by existing security camera networks in large public venues such as airports and transportation hubs to provide physical security personnel with more usable video monitoring technology. Towards this goal, the Department of Homeland Security commenced working with NIST and a number of interested U.S. Government agencies to develop a challenge problem approach to the evaluation of real-time video tracking analytic algorithms that are robust to the environmental variations and camera coverage variability that are characteristic of such venues. The initiative was named Video Tracking Analytics for Physical Security (VTAPS). Three of VTAPS' strategic goals directly relate to standards development:

- designing and implementing video tracking analytic technology challenge evaluations employing realistic large-scale data collections;
- creating scalable re-usable test bed architectures that support both formal evaluations and R & D; and
- developing consensus standards that will facilitate and accelerate technology development and transition.

An example of a VTAPS task is given as follows.

Task: Generate the geospatial path of a given individual (or vehicle or object). Identify cameras the individual appeared in.

Environment: Large busy space such as an airport, transportation centre, city centre, ...

Input: (1) Video streams from a large network of cameras with varying overlapping fields of view and associated camera and environmental metadata, and (2) still or video imagery seed.

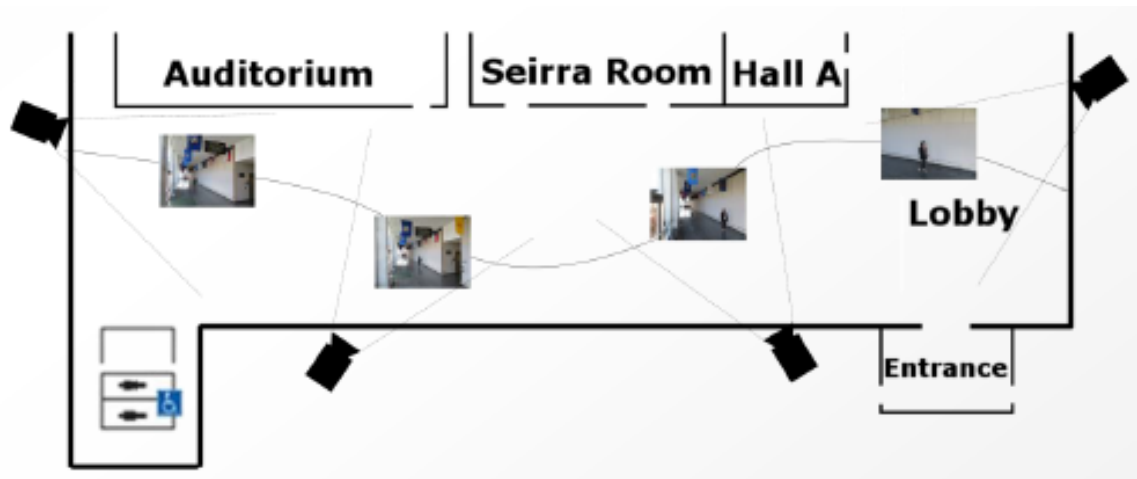


Figure 6: Example of VTAPS challenge

As of June 2016, to the best of the author's knowledge, the VTAPS initiative has not yet resulted in a tracking challenge being set and/or executed. This has largely been due to U.S. Government budgetary constraints since the initiative was launched.

3.1.4.1.14. VAPS (US)

In February 2016, NIST, the OSTP NITRD Video and Image Analytics (VIA) Interagency WG, and the DHS-led Video Quality in Public Safety (VQiPS) WG joined forces to foster the creation of a technically diverse Video Analytics in Public Safety (VAPS) community of interest (CoI) to develop a national R & D strategy in this emerging area and begin critical collaboration, R & D, measurement, and standards activities. The CoI will formally convene in June 2016. Specifically in relation to standards, the group is expected to address best practices and open standards related to the implementation of VAPS. The first stage of the work is focused on identifying the challenges, gaps, opportunities and needs in relation to standards. The strategic priorities identified by the initial VAPS effort will be mapped into the 'Interagency Strategy for Research and Development in Visible World Image and Video Analytic Technologies' being developed by NITRD VIA as well as to the NIST PSCR Analytics Roadmap and future DHS VQiPS collaboration and best practices activities.

3.1.4.1.15. IARPA DIVA (US)

In July 2016, the U.S. Intelligent Advanced Research Projects Activity (IARPA) organised a Proposers' Conference Day for the Deep Intermodal Video Analytics (DIVA) programme (<https://www.iarpa.gov/index.php/research-programs/diva>) ahead of an official announcement. The programme intends to 'develop robust automated activity detection for a multi-camera streaming video environment. As an essential aspect of DIVA, activities will be enriched by person and object detection, as well as recognition at multiple levels of granularity'. Furthermore, the DIVA programme is expected to 'produce a common framework and software prototype for activity detection, person/object detection and recognition across a multi-camera network. The impact will be the development of tools for forensic analysis, as well as real-time alerting for user defined threat scenarios'.

3.1.4.2. Projects

A number of European funded projects have addressed standards and certification as part of their work plan. While the following list is not intended to be exhaustive, it provides a representative set of recent projects, selected based on their relevance to standards in video surveillance and video analytics, their focus on investigative video analysis, and/or their application of existing surveillance standards in real practice.

3.1.4.2.1. Subito

The Subito (www.cordis.europa.eu/project/rcn/89391_en.html) FP7 project addressed the development of automated real-time (CCTV) detection of abandoned luggage and fast identification and tracking of the owner (pre- and post-event). Typical public space surveillance scenes were considered (e.g. car park, exhibition hall, etc.). Supporting studies examined additional sensors. The project developed the metadata language and interoperability required to integrate the surveillance components (detection, tracking, threat detection) and human-computer interface.

3.1.4.2.2. Protectrail

The Railway-Industry Partnership for Integrated Security of Rail Transport (Protectrail) (www.protectrail.eu) was a EUR 21 million, 29-partner project funded by the EU from September 2010 to February 2014 aimed to make rail transport more secure while ensure it remains open, accessible and enables efficient flow of people and goods. Furthermore, to achieve a level of standardisation achieved in other industries. To achieve this, the project developed an interoperable and modular framework consisting of a set of rules and standards, which facilitate the integration and communication amongst various security technologies. It is based on three key ideas: (1) interoperability is improved through standardisation, (2) re-use of existing and relevant international standards is preferred, and (3) simplicity is key to long-term adoption. This includes video management systems as one of the security technologies. An important outcome of the project was that the interoperability framework was tested during field demonstrations concentrating on four priority facets including one on video management. Overall, the conclusion reached was that railway security will be enhanced if 'the multitude of actors in the field will adopt international interoperability standards for security' (Protectrail 2014) (UKGOV 2015).

3.1.4.2.3. SECUR-ED

Secured Urban Transportation — European Demonstration (SECUR-ED) (www.secur-ed.eu) was a 43-partner EUR 40 million project between April 2011 and September 2014. It engaged major operators and top industrial integrators to enhance the security of urban public transportation in medium and large cities, through a series of live demonstrations. SECUR-ED developed a fixed and on-board consistent CCTV and communication architecture to improve security on trams and trains, as well as in depots and stations. In order to make this architecture open, interoperable and future-proof, SECUR-ED relied on two international standards: ISO 22311 and IEC 62676-2. These have been used to define minimum requirements applicable to the different CCTV systems and to form the basis for the development of architectures that were deployed in the demonstrations. There is a strong connection to Protectrail and to reach similar conclusions with regards to limitations in existing standards (see below). It was further noted that in urban surveillance (SECUR-ED 2014):

- there is a need (by authorities) 'to go back to the origin of the incident and then navigate with agility (faster than real time, backwards, etc.) to understand the sequence of events or to perform special tasks, like tracking an individual over the full extent of a city. Tools do not exist today to achieve this level of interoperability between heterogeneous systems and the only option is to extend the scope of existing standards; roadmaps have in fact been put in place to build on IEC 62676 and ISO 22311';

- 'videos are collected to be usable for forensic analysis. This implies minimum video quality (sometimes mandated by law), proper and unambiguous identification of the scenes, time of occurrence and the ability to be decoded by police systems. The abovementioned ISO 22311, recently promulgated, addresses these requirements'.

3.1.4.2.4. ARENA

The EU project ARENA (May 2011-May 2014) (www.foi.se/arena) addressed the design of a flexible surveillance system for detection and recognition of threats towards deployment on mobile critical assets such as trucks, trains, vessels and oil rigs. The objective of ARENA was to develop methods for automatic detection and recognition of threats, based on multisensory data analysis. The ARENA system incorporates all levels from low-level sensor data processing to high-level decision support and HMI-interfaces. ARENA. ARENA contributed to standards developments in three ways (ARENA 2014):

- (1) surveillance architecture: development of a common integration platform architecture (Hołubowicz, et al. 2012) and common data model;
- (2) contribution to wider discussion on standards: as part of the NIST VTAPS workshop on video tracking evaluation, datasets and metrics held in July 2013;
- (3) benchmark development: contribute to establishing surveillance detection, tracking and threat analysis standards by developing appropriate metrics as well as carrying out the collection and annotation of an appropriate dataset that can be employed as a benchmark (PETS2014).

3.1.4.2.5. Savasa

The Standards Based Approach to Video Archive Search and Analysis (Savasa) (www.savasa.eu) was an EU funded initiative, which ran from December 2011 to May 2014. The 12 partner, EUR 4 million project, proposed the 'creation of a video archive search platform that allows authorised users to perform semantic queries over different, remote and non-interoperable video archives'. There was a strong focus on interoperability and open standards (Savasa n.d.) (Savasa 2014). In conclusion, the project found a particular lack of standardisation in surveillance video encryption and surveillance video watermarking.

3.1.4.2.6. Advise

The Advanced Video Surveillance Archives Search Engine for Security Applications (Advise) (www.advise-project.eu) aimed to design and develop a unification framework for surveillance-footage archive systems, specifically to assist law enforcement authorities fight against crime and terrorism via efficient evidence mining into heterogeneous video archives. The EUR 4 million, 11-partner project which ran from March 2012 to February 2015 included a deliverable on standardisation which detailed work in two phases. The first phase made a study on standards relevant to Advise and identification of suitable standards for the Advise system. The second phase performed an evaluation of selected standards and identified potential extensions (see below).

3.1.4.2.7. Forensor

The FOREnsic evidence gathering autonomous sensor (Forensor) EU project (www.forensor-project.eu) which started in September 2015 and runs through to September 2018, is a 11-partner EUR 5 million project tasked with developing and validating a novel, ultra-low-power, intelligent, miniaturised, low-cost, wireless, autonomous sensor ('Forensor') for evidence gathering. The project does not aim to produce new standards, but to actively contribute to ONVIA and PSIA.

3.1.4.2.8. ITEA 2 LINDO

The Large scale distributed INDexation of multimedia Objects (LINDO) (ITEA 2 2010) was an ITEA 2 funded project between November 2007 and October 2010. The project

'demonstrated an effective open system for indexing and retrieving specific objects in very large distributed multimedia archives with remote selection and processing. A few seconds of critical video can be retrieved from thousands of hours of recordings based on any type of criteria. The system offers an integrated solution optimised for video over Internet, implementation of a practical querying mechanism and standardised data formats'. A particular innovation of the work was an agreement of a common data format for video surveillance and links to the work undertaken by ANFOR and subsequently under ISO 223 WG5 (ITEA 2 2010).

3.1.4.2.9. Reveal

The Reveal project (www.computing.surrey.ac.uk/ai/reveal/) was funded by UK's Engineering and Physical Sciences Research Council between September 2004 and May 2008 and addressed the recovery of evidence from video by fusing video evidence thesaurus and video metadata. The strategic objective of the project was to promote those key technologies which enabled automated extraction of evidence from CCTV archives, and to allow integration within the Home Office Large Major Enquiry System crime management system which existed at the time. The project work included development of methods to extract different levels of metadata: *camera-specific*, *scene-specific* and *object-specific* metadata that are extracted using layered image processing algorithms. These metadata streams were then integrated to establish a visual knowledge ontology (or surveillance metadata model) that could extrapolate from low-level computer vision concepts to high-level representations of activity or behaviour.

3.1.4.2.10. Caretaker

The aim of the EU project Caretaker, which ran from March 2006 to August 2008, was to study, develop and assess multimedia knowledge-based content analysis, knowledge extraction components, and metadata management subsystems in the context of automated situation awareness, diagnosis and decision support. Specifically, the project work focused on the extraction of a structured knowledge from large multimedia collections recorded over networks of camera and microphones deployed in real sites (<http://www.multitel.be/image/research-development/research-projects/caretaker.php>). In relation to standards, the project produced a specification of an ontology that represented user and scene knowledge. This represented a standard by which a surveillance database can be queried, and was flexible in order to accommodate new scenarios. Also, as part of the work, annotated datasets were produced for one particular scenario on train stop detection.

3.1.4.2.11. CRISP

'The Evaluation and Certification Schemes for Security Products (CRISP) is a 3-year project (April 2014-March 2017) that aims to facilitate a harmonised playing field for the European security industry by developing a robust methodology for security product certification' (www.crispproject.eu). At the time of writing (June 2016), the project has produced a range of deliverables including a glossary and taxonomy of security products, systems and services, a report on security standards, certification and accreditation — best practices and lessons learnt, and details on evaluation and certification methodology to be applied during the remainder of the project. The final output of the project in 2017 is expected to be an EU Security Certification Manual, specifying standards and requirements for certification and accreditation of security products. The manual will be targeted at certification and accreditation bodies and *inter alia*, set out their roles and responsibilities.

3.1.4.3. Representative benchmark datasets

The following provides a summary of representative video surveillance datasets, which have been extensively used by the community. Some of the datasets (Caviar, Etiseo, PETS) are more focussed on visual tracking, while others (VIRAT) are focused on event recognition. A more detailed description of related datasets is presented in the published ERNCIP TG-VS report on 'Access to data sets'.

3.1.4.3.1. Caviar

The Context Aware Vision using Image-based Active Recognition (Caviar) project datasets include two sets of video clips filmed at separate locations; the first being a building's entrance lobby and the second an indoor city shopping centre. The footage was shot to address analysis of city centre surveillance, both from view of antisocial behaviour and that of potential customers in a commercial setting. The Caviar benchmark datasets collectively show people walking alone, meeting with others, window shopping, and entering and exiting shops, fighting, passing out and leaving a package in a public place. The shopping centre dataset includes two separate viewpoints that are time synchronised.

3.1.4.3.2. Etiseo

The collective datasets of Project Etiseo consist of indoor and outdoor scenes, corridors, streets, building entries, a subway station and an airport apron. For some scenarios, the researchers providing the available datasets recognised that in addition to multiple cameras it is entirely possible that the use of multiple image modalities may bring further benefits towards developing robust solutions. The Etiseo project (described above) presents many of its scenes as multi-camera datasets and some include additional imaging modality such as infrared footage.

3.1.4.3.3. PETS

Since 2000, PETS has collected and disseminated a range of datasets to the scientific community. These include public space surveillance (for example, PETS2000 and PETS2001 dataset on car park and pedestrian surveillance, PETS 2006 on train station monitoring, specifically detection of left (abandoned) luggage, and PETS2009 on crowd image analysis. The most recent datasets produced are for the 2016 PETS workshop and focus on (1) a multi-sensor land case, which addresses the protection of trucks (the ARENA dataset), and (2) a multi-sensor maritime dataset, which addresses the protection of a vessel at sea from piracy. The latter dataset is unique as it comprises a suite of heterogeneous sensors (GPS, visual and thermal cameras) and fills the current void of publicly available event detection and behaviour understanding data in this area.



Figure 7: PETS2009 crowd image analysis dataset

3.1.4.3.4. iLids

The first datasets released as part of the benchmark consisted of four scenarios including sterile zone or perimeter monitoring, parked vehicle detection, abandoned baggage detection and doorway surveillance and monitoring. The dataset was later enhanced in November 2008 to include a fifth multi-camera (person) tracking scenario within a terminal building at Gatwick airport, UK (and which was subsequently made available to the TRECvid project). However, this scenario did not achieve such a high level of interest from the community, at least in terms of product certification requests. Later (Sage, Nilski and Sillett 2010), a sixth dataset series was added which addressed several technology areas: (1) thermal imaging systems, and (2) systems that rely on active IR illumination. Overall, while the published datasets became useful for evaluating detection

algorithms it remained limited because some parts of the dataset are monocular and it also does not contain examples of specific behavioural interactions.

3.1.4.3.5. VIRAT

VIRAT was set up as a large-scale video dataset in 2011 (Oh, et al. 2011) (Dufour 2012) to assess the performance of diverse visual event recognition algorithms with a focus on continuous visual event recognition in outdoor areas with wide coverage. The data consists of multiple outdoor scenes with actions occurring naturally by non-actors in continuously captured videos of the real world. The dataset includes large numbers of instances for 23 event types distributed throughout 29 hours of video. This data is accompanied by detailed annotations that include both moving object tracks and event examples, which provide a solid basis for large-scale evaluation. Additionally, different types of evaluation modes for visual recognition tasks are proposed as well as a set of evaluation metrics.

3.1.4.4. Metrics

The measurement of the performance of video analytic algorithms (for example, tracking or event detection) is necessary to quantify how reliable an algorithm is in a particular surveillance scene. Many types of metrics have been proposed over the years to address this issue and most of them are dependent on ground truth data in order to compare analytic output with the ideal. While many evaluation initiatives have used CLEAR and VACE metrics, new metrics continue to be proposed due to lack of consensus on standardised ways of assessing performance of tasks such as object tracking. Examples of recent work in the literature include Nawaz, Poiesi and Cavallaro (2014).

3.1.4.5. Ground truth

Ground truth (or labelled data) is a crucial component in the evaluation of the performance of video surveillance algorithms. Depending upon the particular task to be evaluated (for example, detection, tracking or event recognition), benchmark data, as described under Section 3.1.4.3, is labelled. Ground truth refers to the ideal performance an algorithm is desired to achieve. High quality labels enable the comparison of processed image output to be compared against the ground truth (method evaluation), and to construct statistical models. It is commonly recognised that one of the most tedious, time consuming and error-prone aspects when developing benchmarks, is the generation of ground truth. An active area of research is to generate a tool that enables fast ground truth generation.

3.1.5. Certification of surveillance systems

Areas such as detection equipment for aviation security and IT security are highly regulated. However, this is not generally the case for surveillance systems (Lewis 2014). Certain parts of the system are well covered. For example, CCTV and video technology are well covered by the latest IEC 62676 set of standards (Lewis 2014) and in the UK there exists a national standard for automatic number plate recognition (ANPR) systems which must comply with the standard to be connected to the national ANPR data centre (CRISP 2014). Certification is provided by a range of bodies. As detailed under Section 3.1.4.1.1, BRE Global Ltd are one such provider who currently certify video analytics algorithms/systems applied to the iLids dataset. A very important gap identified by the ERNCIP TG-VS, is that there is no European certification process for the performance of video surveillance (analytic) systems, and there exists a lack of known standards to certify to.

3.2. Gap analysis

3.2.1. Interoperability

The forensic video surveillance investigative case study, described under Section 2.1, identified a number of areas where there exists a lack of standards. This includes:

- Level 1 — Technical interoperability: storage and transmission
- Level 2 — Syntactic interoperability: video formats and protocols
- Level 3 — Semantic interoperability: video analytics and manual annotation
- Level 4 — Pragmatic interoperability: how algorithm, methods and procedures are applied.

Recent standards, including ISO 22311 and NF EN 62676 have made some headway in this direction, especially in lower levels of interoperability. However, much remains to be done to address the full chain of processing for a given video surveillance application. For example, in forensic video investigations, there does not exist an interoperable video archive search platform. If such a platform existed this would significantly enhance forensic capability across the EU.

3.2.1. Certification

There does not exist any European level certification for surveillance systems or its components. More generally, there does not exist a design process and methodology for testing and validation of surveillance solutions (Gemo and Andritsos 2011). An inhibitor to progress in this area is the lack of a procurement framework, which requires and builds upon certification (J. van Rest, Surveillance use cases: Focus on video analytics 2015). Considering video surveillance architectures as an example, in many surveillance works new architectures are developed which duplicates effort and makes integration more complex. This demands development of a reference architecture implementation: a trusted, secure, scalable, extensible and interoperable open source architecture based on open standards and incorporating privacy-by-design which could be applied to a wide range of surveillance applications (for example, public space surveillance, critical infrastructure protection, counter piracy, etc.). Privacy protecting features would be embedded at all levels.

3.2.2. Benchmarking methodology

As described above, standards also lack in a universally agreed set of performance evaluation benchmarking datasets and metrics for video analytics.

3.2.3. EU projects

Further gap analysis can be established through experience of implementation of real operational systems, representative examples of which are described in the following subsections.

3.2.3.1. Protectrail

The Protectrail EU project made a significant impact in the implementation of video surveillance standards (specifically IEC 62676-1&2 and ONVIF profiles) in the real world. However, it was determined that such standards could not be implemented standalone and need to be part of a wider implementation remit including stability, trustworthiness and regulatory (for example, privacy protection) aspects.

Protectrail on its completion in 2014 made the following recommendations with regard to video streams, including quality of service and potential cyber attacks:

- a generalised use of RTP/RTSP streams carrying video H264 compressed metadata time stamped at the frame level;
- full modularity of the basic services associated to video, independently of their physical implementation;
- video surveillance systems are networks of distributed PCs; as such they are potential targets of cyber-attacks, against which they must be protected (physically, by training staff or with software);
- digital video, especially when live information with low latency is required, has stringent needs for communications channels (no buffering is allowed); this

implies a good quality of service for the communication, but also an optimised set-up in the network architecture to minimise throughput at any point of the network in all circumstances (typically a case by case trade-off between Unicast and Multicast);

- the system must preserve full consistency between time and metadata associated with the streams, the events produced by video analytics and the supervision tools. By law the operators generally cannot access the recorded video files for privacy protection reasons. If the operators want to use the video for operational security or training, they have to remove the privacy related attributes for instance by using face blurring. If the control centre requires access to the on-board videos in real time, the infrastructure is not capable today of obtaining constant video streams. ONVIF and RTSP are made for networks with a constant bitrate. For videos streaming on wireless networks the solution is an adaptive bitrate for video streaming depending on the existing wireless infrastructure.

In addition, the following recommendations were made in relation to video analytics:

- a minimum configuration required for analytics: For example, many analytics require an initial calibration for each camera (e.g. to determine its 3D location and orientation or to adjust to internal lighting conditions). To make larger setups (50+ cameras) manageable it is recommended to either automate these calibration procedures with sufficient quality or to use solutions that do not require such configuration;
- using analytics for decision support and not as fully automated security solutions: Complex systems are never 100 % fail safe or fail in unexpected conditions. An interactive system provides functionality to support an operator who is the human-in-the-loop;
- metadata standardisation: Full consistency for video analytics remains an open issue as there are no well-established industry standards and video analytics are a quickly evolving market. Maintaining consistent metadata definitions will require attention when solutions are integrated, especially when a new solution needs to fit into a legacy system. In this situation, the most future-proof approach is to stick on the minimum criteria for events outlined above and rely on associated URLs for details;
- wider system (e.g. storage/playback/GUI/other) requirements: Video analytics usually need more performance or have wider requirements than basic video solutions. Some analytics require for instance high frame rates/high resolution playback of stored data instead of a lower resolution, lower frame rate data. Other analytics might need an extra monitor because they require certain user interactions or provide information that cannot be displayed on a video stream. It is recommended that the video surveillance systems that might be extended at a later time with analytics are designed for upgrade, typically to support analytics (e.g. room for servers or extra monitors, etc.).

Protectrail also stated that collected videos must be usable for forensic analysis. This implies minimum video quality (sometimes mandated by law), proper and unambiguous identification of the scenes, time of occurrence and the ability to be decoded by police systems. ISO 22311 addresses these requirements.

Finally, Protectrail also recognised that video surveillance can be extremely useful for security management and crime investigation, but that it also might result in an unnecessary intrusion into citizen privacy. When video surveillance is used as a balanced guide by regulations complemented by common sense, it needs to be struck.

3.2.3.2. Savasa

The Savasa EU project also determined that a standard for video encryption and standard for surveillance video watermarking, which would be recognised by legal authorities, had not yet been identified.

3.2.3.3. Advise

The Advise EU project identified the following potential extensions in standards (Advise 2012):

- MPEG-7 standard: 'it appears MPEG-7 standard could be suitable for representation of the final results from the Advise system that are extracted for use in a court of law. The MPEG-7 standard is about creating and then searching metadata of particular relevance to security video. Since the goal in video surveillance is to call out and identify suspicious behaviours and security breaches or events, details and fine recognition in viewing isn't emphasised, though available as needed. Because MPEG-7 inherently searches for preidentified behaviours on which to focus, it is well suited for surveillance security analytics';
- MPEG-A standard: 'the standard provides wide range of interesting and suitable solutions for managing interoperability in video surveillance systems'.

4. Basic elements for standardisation process concerning use of video surveillance systems

4.1. Standardisation process

The main questions that need to be answered when considering development of new standards (in video surveillance or any other area) are:

- Is it a product, service or process that needs to be standardised?
- Why is the standard necessary?
- For which industry or industries is it relevant and specifically, who would use the standard?
- Are there currently any relevant standards or is there any relevant existing good practice within the industry?

The process starts by establishing the need for a standard and to bring together all interested parties to share what they need, what they have, and details on what to include in the standard.

For development of both EU (BSI 2011) (Advise 2012) and international (ISO) standards, experts are convened via national standardisation bodies and convened to a single TC. A draft standard is jointly developed and national members carry out a public enquiry involving all interested stakeholders. The comments are evaluated and the standards draft adapted as needed. National members then vote on the consensus draft and finally the standard is published. The standard may be sent back to the TC for further modification as necessary.

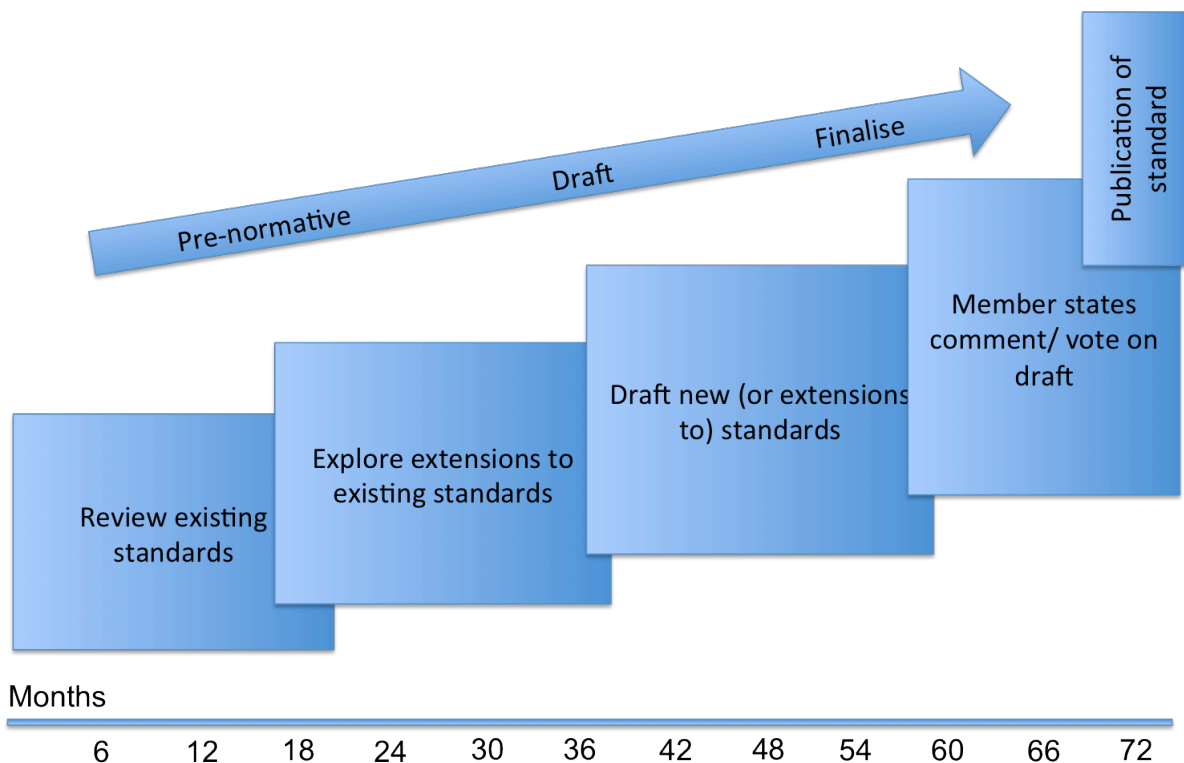


Figure 8: Key stages in development of new standards

Overall, the standardisation process takes a considerable amount of effort and elapsed time before a new standard is produced. The pre-normative work required to develop a new standard would be expected to take a minimum of 3 years and includes reviewing existing standards and exploring extensions to those standards. A further 3 years is then required before the new standard is first published. Hence, the overall process would normally be expected to take 6 years. During this time, it can be expected that a number of intermediate results would be obtained, including informal specifications of the new standard, *de facto* standards, and reference systems.

4.2. Recommendations for video surveillance standards

Based on the findings in this report, the main recommendations to be made are as follows:

To action a new work item to develop one or more EU standards for surveillance of critical infrastructure.

The goal is to define a standard for video surveillance for protection of critical infrastructure. The standard will enable efficient use of video surveillance data by addressing the current technical interoperability problems on multiple levels, as described in Section 2.1. More specifically, it will address the lack of interoperability in existing surveillance systems by analyzing existing constraints and designing, implementing and evaluating new and efficient video surveillance solutions.

The following **pre-normative** work steps will need to be completed in order to initiate a new work item:

- (1) analysis of the constraints on technical interoperability given by state-of-the-art hardware, and investigate current hardware developments to be able to point out trends in, for example, dedicated hardware for surveillance of critical infrastructure;
- (2) address the need for syntactical interoperability, i.e., video formats and protocols by analyzing the European CCTV market and build upon existing technology for accessing CCTV imagery from a multitude of systems;

- (3) address the need for semantic interoperability and enable the exchange of meaningful information between different CCTV systems. This includes development of appropriate metadata standards which in turn could be achieved through closer cooperation between the ERNCIP TG-VS and ONVIF and PISA;
- (4) address the need for pragmatic interoperability, describe the circumstances in which a particular annotation is strong, and automatically determine the context of the recordings of the video signal.

It will also be necessary to determine whether to pursue the CEN or ISO route for new standards. The CEN route would be the default for the ERNCIP TG-VS however, it does not preclude ISO if there already exists relevant work underway or identified countries that have shown interest to achieve the same goal. The next step would be for the ERNCIP TG-VS to consult with CEN directly through the pre-existing contact point that the TG-VS has established.

To action a new work item to develop a harmonised certification procedure for video surveillance systems and components for protection of critical infrastructure at EU level.

This should include research in, and community agreement on, performance evaluation methodology (datasets and metrics) for video surveillance/analytic application to protection of critical infrastructure. The ambition is to develop an EU compliant open certification standard for a wide range of video surveillance tasks, based on common configurations. This could form the basis of a European Test and Validation Centre. Such certification would support a procurement framework to be used by critical infrastructure end users when procuring video analytics making use of these datasets (J. van Rest, Surveillance use cases: Focus on video analytics 2015).

5. Roadmap to achieve draft standard agreement concerning new standards in, and certification of, video surveillance systems

The standardisation stages shown in Figure 8 are similar for the development (or revision) both ISO and CEN standards. The aim under ISO is to reach the Committee Draft stage, and under CEN it involves obtaining a Workshop Agreement. The key standardisation stages for the video surveillance work items identified in Section 3.2, and which map onto the general ISO/IEC stages shown in Figure 8, are respectively detailed in the following sections.

5.1. New standards in surveillance of critical infrastructure

- Produce a detailed evaluation of the most relevant existing standards for the video surveillance of critical infrastructure.
- Explore potential extensions or proposed modifications to the identified existing standards to fulfill the surveillance needs, based on current best practice and operational procedures in surveillance and related domains.
- Draft extension/modification to existing standards or for new standard(s) with support of all relevant stakeholders.
- The countries (ISO members) that have chosen to participate in the development of the standard(s) form a national position on the draft and comment on it.

5.2. Certification of surveillance systems for protection of critical infrastructure

- Undertake a common requirements gathering exercise to collect the views and opinions of relevant surveillance system stakeholders on the possibility of creating a harmonised certification procedure for video surveillance systems and components at EU level.

- Based on positive support established in the consultation with stakeholders, produce an impact assessment in this sector.
- Undertake a security profiling of products to be certified.
- Research and develop an appropriate video surveillance system evaluation methodology, including appropriate dataset(s) and metrics.
- Draft 'new approach' guidance for the sector, i.e. with technical requirements defined in harmonised standards and with obligatory conformity assessment performed by 'notified bodies'; organisations selected for the task by the Member States and notified to the Commission by them.
- Certification would support a procurement framework to be used by critical infrastructure end users when procuring video analytics making use of these datasets (J. van Rest, Surveillance use cases: Focus on video analytics 2015).

6. Conclusions

It is widely recognised that standards play a major role as fundamental building blocks in surveillance product development, to ensure uniform quality in provision of services, and wider still in enabling the EU security industry to be more competitive globally.

However, there exist very few standards in the security domain, and in particular, in video surveillance systems.

This document has provided an introduction to standards in video surveillance, including the need for standards, an overview of existing relevant standardisation efforts including gaps, and recommendations and roadmap for future standards development.

Overall, the aim is to promote the exchange of good and best practices for application of video surveillance for critical infrastructure protection, and assist in the development of a single market in the EU for critical infrastructure protection related products and services.

6.1. Next steps for TG-VS

In summary, the following concrete actions by TG-VS would further standards' development in video analytics and surveillance for critical infrastructure protection:

- to develop a procurement framework to be used by critical infrastructure end users when procuring video analytics;
- to undertake preparatory work in the development of one or more new standards in video surveillance (see Section 5.1 above);
- to undertake preparatory (pre-normative) work in the certification of surveillance systems (see Section 5.2 above).

The main audience for the recommendations and next steps are EU level policy authorities, specifically the Directorate-General for Migration and Home Affairs. The audience for the report as a whole further include industry, end users and other stakeholders with interest in deployment of video analytic and surveillance systems and methods, and academics researching surveillance and standardisation.

References

How many video surveillance cameras are there in this world?
<https://storageservers.wordpress.com/2014/07/30/how-many-video-surveillance-cameras-are-there-in-this-world/> (accessed 30 July, 2014).

Advise, *D2.1 Recommendations for interoperable standards and existing solutions for video archive search*, 2012.

Advise, *D8.16 Report on standardisation*, 2015.

ARENA, *ARENA WP9: Standardization Report*, 2014.

Bernardin, K. and Stiefelhagen, R., 'Evaluating multiple object tracking performance: The CLEAR MOT metrics', *EURASIP Journal on Image and Video Processing*, 2008.

BSI, *A standard for standards — Principles of standardization*, 2011.

CRISP, *Deliverable D2.2: Consolidated report on security standards, certification and accreditation — best practice and lessons learnt*, CRISP EU project, 2014.

Dufour, J.-Y., *Intelligent Video Surveillance Systems*, edited by Dufour, J.-Y., 2012.

European Commission, *M/487: Programming mandate addressed to CEN, CENELEC and ETSI to establish security standards*, Enterprise and Industry Directorate-General, 2011.

Gianosvaldo, F. and Umiliacchi, P., *New standard for multimedia applications on board trains enables improved interoperability*, 9th World Congress on Railway Research, 2011, http://www.railway-research.org/IMG/pdf/poster_fadin_gianosvaldo.pdf

Wen, G., Tian, Y., Huang, T., Ma, S. and Zhang, X., 'The IEEE 1857 standard: Empowering smart video surveillance systems', *IEEE Intelligent Systems*, Vol. 29, No 5, September 2013, pp. 30-39.

Gemo, F., and Andritsos, F., *Workshop on emerging surveillance capabilities and requirements*, JRC Ispra, 2011.

Gorodnichy, D. O., Laganier, R. and Macrini, D., *Video analytics evaluation: survey of datasets, performance metrics and approaches*, Government of Canada, 2014, <http://pubs.drddc-rddc.gc.ca/BASIS/pcandid/www/engpub/DDW?W %3DSYSNUM=800521>

Hołubowicz, W., Knapik, R., Samp, K., Szklarski, L. and Taberski, G., *Architecture of a surveillance system for detection and recognition of threats for deployment on critical mobile assets/platforms such as vessels, oil rigs, trucks and trains (on the basis of the 7FP ARENA project)*, 5th International Scientific-Technical Conference (Natcon), 2012.

Home Office, *CCTV operational requirements manual 2009*, 2009.

IEEE Computer Society, *1857a-2014 — IEEE standard for advanced audio and video coding — Amendment 1: Extension on timing and location information to support object tracking across multiple cameras at surveillance high group*, April 2014, pp. 1-34.

ISO, 'ISO 22311 Societal security — Videosurveillance — Export interoperability', *ISO TC 292*, <http://www.isotc292online.org/projects/ISO22311-rev/> (accessed February 29, 2016).

ITEA 2, *LINDO innovation report*, ITEA 2, 2010, <https://itea3.org/project/result/download/5752>

ITEA 2, *LINDO: Large scale distributed INDEXation of multimedia objects*, 2010, <https://itea3.org/project/lindo.html>

Kasturi, R. et al., 'Framework for performance evaluation of face, text, and vehicle detection and tracking in video: Data, metrics, and protocol', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 31, No 2, February 2009, pp. 319-336.

Leal-Taixé, L., Milan, A., Reid, I., Roth, S. and Schindler, K., *MOTChallenge 2015: towards a benchmark for multi-target tracking*, 2015, arXiv.org

Lewis, A. M., *Technology certification for critical infrastructure protection: Current procedures in EU and EEA States*, European Commission, 2014.

Marcenaro, L., *Access to datasets*, JRC, 2016.

Nawaz, T., Poiesi, F. and Cavallaro A., 'Measures of effective video tracking', *IEEE Transactions on Image Processing*, Vol.23, No 1, January 2014, pp. 376-388.

Sangmin, O. et al., 'A large-scale benchmark dataset for event recognition in surveillance video', *2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2011, pp. 3153-3160.

Poustourli, A. and Kourti, N., 'Standards for critical infrastructure protection (CIP) — the contribution of ERNCIP', *HE EURAS Board Series (EURAS Contributions to Standardisation Research)*, 2014, <http://publications.jrc.ec.europa.eu/repository/handle/JRC91182>

Protectrail, *Protectrail white paper: Key lessons for the railway sector on the Protectrail security architecture*, 2014, http://www.protectrail.eu/IMG/pdf/protectrail-white_paper-final.pdf

Rest, J. H. C. van, *Terminologie en taxonomie van video content analyse (Terminology and taxonomy of video content analysis)*, TNO, 2010.

Sage, K., Nilski, A. and Sillett, I., 'Latest developments in the iLids performance standard: New imaging modalities', *EEE Aerospace and Electronic Systems Magazine*, 2010, pp. 4-11.

Sapori, E., Sciutto, M. and Sciutto, G., *A quantitative approach to risk management in critical infrastructures*, 17th Meeting of the EURO Working Group on Transportation (EWGT2014), Seville, 2014.

Savasa, *D5.41: Good practices for data protection and privacy*, 2014, http://savasa.feedribbon.es/remote/fr_documents/magazines/SAVASA_D5.41_Good_practices_for_Data_Protection_and_Privacy_v1.0.pdf (accessed February 29, 2015).

Savasa, *Savasa public deliverables*, http://savasa.feedribbon.es/remote/fr_documents/magazines/SAVASA_D5.51_Model_migration_path_v1.0.pdf

Savasa, *Savasa public deliverables*, 2014, http://savasa.feedribbon.es/remote/fr_documents/magazines/SAVASA_D5.61_Proposal_and-or_colaboration_to_standardization_v1.0.pdf

Schallauer, P., Bailer, W., Hofmann, A. and Mörzinger, R., *SAM: An interoperable metadata model for multimodal surveillance applications*, SPIE 2344, 2009.

SECUR-ED, *D46.4 Report on consolidation of the interoperability level achieved*, 2014.

SECUR-ED, *SECUR-ED white paper on public transport stakeholders: Based on the lessons learned in SECUR-ED*, http://www.secur-ed.eu/wp-content/uploads/2014/11/SECUR-ED_White_Paper_Draft3.pdf

Sulzer, J.-F., 'Video-surveillance standardization and expected operational benefits', *on Emerging Surveillance Capabilities and Requirements*, 2014.

Tolk, A., Diallo, S. and Turnitsa, C., 'Applying the levels of conceptual interoperability model in support of integratability, interoperability, and composability for system-of-

systems engineering', *Systemics, Cybernetics and Informatics*, Vol. 5, No 5, 2007, pp. 65-74.

UKGOV, *Recommended standards for the CCTV industry*, Surveillance Camera Commissioner, 2015, <https://www.gov.uk/guidance/recommended-standards-for-the-cctv-industry>

van Rest, J., *A maturing industry depends on standards in engineering surveillance systems*', Workshop on Emerging Surveillance Capabilities and Requirements, JRC, 2011.

van Rest, J., *Surveillance and video analytics: factors influencing the performance*, JRC Science Hub, 2015.

van Rest, J, *Surveillance use cases: Focus on video analytics*, JRC, 2015.

van Rest, J. et al., 'Requirements for multimedia metadata schemes in surveillance applications for security', *Multimedia Tools and Applications*, Springer, May 2014, pp. 573-598.

Zhang, X., Hang, T., Tian, Y. and Gao, W., 'Overview of the IEEE 1857 surveillance groups', *20th IEEE International Conference on IEEE*, 2013, pp. 1505-1509.

List of abbreviations and definitions

BSI British Standards Institute

Cenelec European Committee for Electrotechnical Standardization

ETSI European Telecommunications Standards Institute

Interoperability The ability of communication of systems and units to provide services and to accept services from other systems and units, in order to use the services for efficient operation. Ability for information or services to be exchanged, directly and smoothly, between providers and consumers.

ISO International Organisation for Standardization

JTC Joint Technical Committee

MPEG Moving Picture Expert Group

NATO North Atlantic Treaty Organization

NIST National Institute of Standards and Technology

ONVIF Open Network Video Interface Forum

PETS Performance Evaluation of Tracking and Surveillance

PSIA Physical Security Interoperability Alliance

SAM Surveillance Application Metadata

Standard Document, established by consensus and approved by a recognised body, that provides, for common and repeated use, rules, guidelines or characteristics for activities or their results, aimed at the achievement of the optimum degree of order in a given context

Standardisation Activity of establishing, with regard to actual or potential problems, provisions for common and repeated use, aimed at the achievement of the optimum degree of order in a given context

TC Technical Committee

TG-VS ERNCIP Thematic Group on Video Surveillance for Security of Critical Infrastructure

VBDS Video Based Detection System

WG (Working Group) Group of experts, appointed by an international or European technical committee or subcommittee, responsible for drafting document

List of figures

Figure 1: Surveillance use case	8
Figure 2: Levels of interoperability	9
Figure 3: Main standards relating to video surveillance for protection of critical infrastructure (adapted from (Sulzer 2014)).....	15
Figure 4: ISO 31000 risk management process	17
Figure 5: Timeline of evaluation initiatives	18
Figure 6: Example of VTAPS challenge.....	23
Figure 7: PETS2009 crowd image analysis dataset	27
Figure 8: Key stages in development of new standard	32

Europe Direct is a service to help you find answers to your questions about the European Union
Free phone number (*): 00 800 6 7 8 9 10 11
(*) Certain mobile telephone operators do not allow access to 00 800 numbers or these calls may be billed.

A great deal of additional information on the European Union is available on the Internet.
It can be accessed through the Europa server <http://europa.eu>

How to obtain EU publications

Our publications are available from EU Bookshop (<http://bookshop.europa.eu>),
where you can place an order with the sales agent of your choice.

The Publications Office has a worldwide network of sales agents.
You can obtain their contact details by sending a fax to (352) 29 29-42758.

JRC Mission

As the Commission's in-house science service, the Joint Research Centre's mission is to provide EU policies with independent, evidence-based scientific and technical support throughout the whole policy cycle.

Working in close cooperation with policy directorates-general, the JRC addresses key societal challenges while stimulating innovation through developing new methods, tools and standards, and sharing its know-how with the Member States, the scientific community and international partners.

*Serving society
Stimulating innovation
Supporting legislation*

