Videosurveillance and privacy: covering the two sides of the mirror with DRM

Juan R. Troncoso-Pastoriza,
Pedro Comesaña
Signal Theory and Communications Department, University of Vigo
ETSI Telecom, Lagoas Marcosende s/n
36310 Vigo, Spain
{troncoso,pcomesan}@gts.tsc.uvigo.es

Luis Pérez-Freire,
Fernando Pérez-González
GRADIANT - Galician R&D Center in Advanced Telecommunications
Gradiant ETSI Telecom, Lagoas Marcosende s/n
36310 Vigo, Spain
{lpfreire,fperez}@gradiant.org

ABSTRACT

Privacy and security have always been key concerns for individuals. They have also been closely related concepts: in order to increase their perception of security, people sacrifice a part of their privacy by accepting to be surveilled by others. The tradeoff between both is usually reasonable and commonly accepted; however, the case of videosurveillance systems has been particularly controversial since their inception, as their benefits are not perceived to compensate for the privacy loss in many cases. The situation has become even worse during the last years with the massive deployment of these systems, which often do not provide satisfactory guarantees for the citizens. This paper proposes a DRM-based framework for videosurveillance to achieve a better balance between both concepts: it protects privacy of the surveilled individuals, whilst giving support to efficient automated surveillance.

Categories and Subject Descriptors

K.4.1 [**Computers and Society**]: Public Policy Issues—*Privacy, Intellectual Property Rights*; D.4.6 [**Operating Systems**]: Security and Protection—*Access Controls, Authentication, Information Flow Controls*

General Terms

Security, Legal Aspects, Human Factors

Keywords

Privacy, security, videosurveillance, access control, rights management

1. INTRODUCTION

Since the beginning of time, any species that makes its way through evolution must have developed mechanisms to ensure its own security and protection. Human kind is no exception to this rule, and thus, security has always been a concern to our species. Our characteristic advantage for achieving this goal resides in technology. When it reached the required maturity level, videosurveillance was a natural step in this direction. The problem comes when security collides with privacy. Probably the best definition of privacy so far was given by Westin in 1970 [37]: *"Privacy is the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others."* Individuals accept to give away a part of their privacy in exchange for greater security, but up to what extent? In terms of privacy, the situation has become worse in the last years due to the rise of IP videosurveillance.

Figure 1 shows the main components of a modern IP videosurveillance network: basically it is composed of a number of IP cameras which may be in very disparate locations but all of them are connected to a control center via an IP network (either local or through the Internet). In the control center, the incoming streams are managed by one or more processing servers, and they can be stored in hard disks for *a posteriori* access. Typically, there is one or more human operators that supervise in a video console the recordings in search for incidents and other relevant events. Moreover, modern videosurveillance systems are beginning to feature functionalities for automated image analysis which make easier the task of the human operators whilst increasing security. However, this increased efficiency in detecting events and collecting information is raising even more privacy concerns. This motivates the need for a joint framework addressing the tradeoff between privacy and security. In this paper, we propose such a framework, and prove that, with slight modifications, the paradigm of DRM can yield a solution that covers both aspects of videosurveillance, namely privacy rights of surveilled people and automation targeted towards security: the two sides of the mirror.

This paper is organized as follows. Section 2 introduces the general framework of privacy in videosurveillance from the point of view of the current European legislation. Section 3 reviews the past works on privacy management and protection, paying special attention to the videosurveillance scenario, and also states the current situation of automated videosurveillance. Section 4 presents our DRM-based proposal, and Section 5 offers a high level perspective on its implementation. The application to a real scenario is illus-

DRM'09, November 9, 2009, Chicago, Illinois, USA.

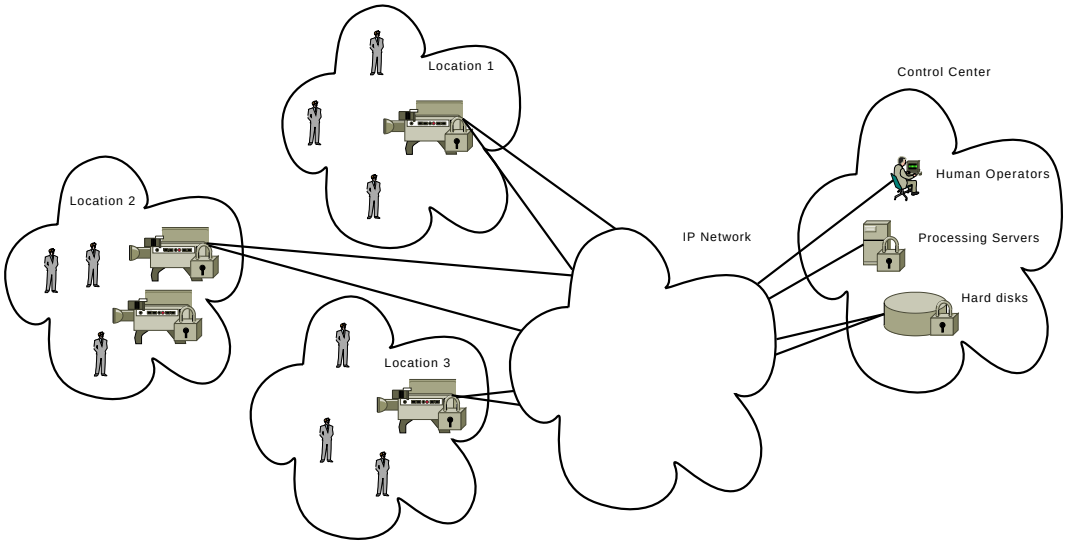

Figure 1: Simplified architecture of an IP videosurveillance network

trated with a use case, described in Section 6. Finally, some concluding remarks are given in Section 7.

2. THE LEGAL FRAMEWORK OF VIDEOSURVEILLANCE

In the last years, the exposition of people to videosurveillance systems has increased considerably. Although these systems are usually perceived as a good means to improve individuals security, as they help to prevent, investigate, detect and prosecute criminal offences, an increasing concern exists about the subsequent reduction of individuals' fundamental rights and freedoms, specially in those aspects related to privacy. This concern was materialized in the International Conference of Data Protection Authorities, held in London during 2006, where *"the need to adapt videosurveillance to the demands of the fundamental right to data protection"* was addressed. In this regard, many different states all over the world with a prior legal framework in privacy are implementing a series of measures aimed at legislating the framework where the activity of the aforementioned videosurveillance systems can be performed, and specifying the necessary conditions such that their activity can be carried out whilst protecting citizens' privacy rights. In this section we will focus on the description of European and Spanish legislations concerning videosurveillance systems, as the European Union (EU) is one of the few domains in the world that currently has a comprehensive set of privacy legislations, with specific legislation for the handling of personal data [12]. As for the Spanish legislation, it represents one of the implementations of the recommendations set forth in the EU. Nevertheless, it must be noted that similar laws rule in most countries.

From a European point of view, Directive 95/46/EC [4] deals with the *"protection of individuals with regard to the processing of personal data and on the free movement of such data"*. This right to privacy is recognized in Article 8 of the European Convention for the Protection of Human Rights and Fundamental Freedoms, promoted by the Council of Europe Convention the 28th January 1981, for the Protection of Individuals with regard to Automatic Processing of Personal Data, and in the general principles of Community Law. It is worth pointing out that, according to this directive, processing of personal data covers *"any operation or set of operations which is performed upon personal data, whether or not by automatic means, such as collection, recording, organisation, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, blocking, erasure or destruction"*. Therefore, given this definition of *"processing of personal data"*, it is straightforward to see that videosurveillance systems are one of the many scenarios covered by the broad scope of this directive. Hence, all videosurveillance systems in the EU should be compliant with it.

According to the principles of protection derived from this Directive, the person/body responsible for processing personal data (the *controller*) must provide to the citizens information about its own identity, the pursued purposes with the processing of their data, and who is the recipient of such data. Furthermore, the citizens are entitled to: 1) know whether or not data relating to them are being processed; 2) know the logic involved in any automatic processing of data concerning them; 3) rectify, erase or block data whose processing is not compliant with the legislation in force. These principles of protection must be applied to any information concerning an identified or identifiable person. In addition, it must be also considered that *"Member States may, in the interest of the data subject or so as to protect the rights and freedoms of others, restrict rights of access and information"*. Furthermore, the Directive states that *"any processing of personal data must be lawful and fair to the individuals concerned whereas, in particular, the data must be adequate, relevant and not excessive in relation to the purposes for which they are processed,"* limiting in this way the personal data that can be manipulated by a videosurveillance system.

Nevertheless, all the previous provisions are subject to exemptions or derogation. For instance, the directive is not

applicable when processing data is intended for journalistic purposes, or when concerning exclusively personal or domestic data, as well as in those activities regarding *"public safety, defence, State security or the activities of the State in the criminal laws."*

Last, but not least, the aforementioned Directive also indicates that Member States should establish supervisory authorities, which must help to ensure transparency of processing in their corresponding countries. In the Spanish case, this role is played by the Spanish Data Protection Agency (AEPD, Agencia Española de Protección de Datos), that in November 2006 proclaimed the Instruction 1/2006 [11], *"on processing personal data for surveillance purposes through camera or video-camera systems,"* which is aimed at ensuring that videosurveillance systems are compliant with the principles of the Organic Act 15/1999 *on Personal Data Protection* [5] (LOPD, according to its Spanish acronym), the national law that materializes the Directive 95/46/EC [4]. Most aspects considered in the LOPD are a direct translation of the principles established in the European Directive to the Spanish videosurveillance framework. Nevertheless, the Instruction specifies in some cases a further detailed implementation. Some of the additional constraints are:

- The controller's information duty binds to *"place at least one informative sign in the areas under videosurveillance, in a sufficiently visible location, in open as well as enclosed spaces*."
- *"The data will be cancelled within the maximum term of 1 month from being gathered."*
- *"The controller must take the measures of technical and organizational nature required to guarantee the security of the data and avoid their alteration, loss and unauthorised processing or access."*

Even when we have centered our attention on the European and Spanish regulations, they are indeed a representative example, and it must be noted that in most countries the legislation in force sets the basis for privacy protection in the videosurveillance scenario.

3. PRIOR ART

Since the late 90's, privacy in the digital world has become a major concern. In this section we review the main initiatives regarding privacy management at a global level and in the particular case of the videosurveillance scenario.

3.1 Privacy management

So far, most of the work in this area has been put in the context of web services. Whereas a bunch of bundled commercial solutions for managing user privacy at the client side are available, solutions addressing privacy issues at a global level are not widely deployed yet.

Kenny and Korba described in [23] a global approach to privacy management that consists in adapting traditional DRM for building an architecture capable of satisfying the European Directive 95/46/EC for privacy protection. The proposed Privacy Rights Management (PRM) architecture is motivated by the inexistence of proper technological means for fulfilling the Directive requirements in web transactions, where users provide private data which is processed by another entity. The authors of [23] propose a client-server architecture with three key elements: 1) the data subject, which is the originator and owner of the private data; 2) the data controller, which is in charge of managing the collection, storage and processing of the private data, and is the ultimate responsible for the misuse of such data; 3) the data processors, which either may depend directly on the data controller or may be third entities.

The data controller manages a system comprising a web server for interfacing with the users, and a PRM server for managing the transactions with data processors and a set of databases. These databases contain the (cryptographically protected) private data of the users, logs on data use, information about the data processors and data subjects, and the rights database which defines the operations that can be carried out on the private data (in other words, the privacy rules). These rights can be conveniently expressed, as for any DRM system, in standard machine readable language using XrML [1], REL [35], etc. Thus, the private data in a PRM system plays the role of "asset" in classical DRM, which is protected according to the legislation in force (the rights). In addition, the data subject could issue a license by which he/she grants certain rights to the data processor in case certain conditions apply.

The authors show in [23] that the proposed PRM architecture fits well the privacy requirements of the Directive 94/45/EC. Howewer, the proposal is not absent of drawbacks. One of the problems is the trustworthiness of the data controller, which processes the private data in the clear. Yet another serious issue, due to its client-server architecture, is scalability, important if a large scale deployment is to be made.

A more recent and popular initiative for privacy management is the Transparent Accountable Datamining Initiative (TAMI) [36], promoted by the MIT, which advances on the basis set by the P3P project [3]. TAMI advocates for transparency and accountability of the private data use, instead of restricting the access to private data. The rationale is that the World Wide Web is making more and more difficult to restrict access to decentralized data, and it is making it easier to aggregate data from multiple information sources, specially for data mining purposes. Hence, it is reasonable to devote efforts to the enforcement of fair data use. However, the problem in videosurveillance is different, as the scenario is bounded and well defined, thus making it easier to implement a DRM-like approach. Moreover, transparency and accountability can be ensured by means of secure event reporting. The TAMI project is particularly focused on large scale data mining. It proposes a "policy aware" architecture comprising rule languages that are able to express policy constraints, and reasoning engines able to produce inferences and proofs for private data use being compliant with the relevant rules.

3.2 Privacy in videosurveillance

Even if societal and ethical privacy concerns due to the rise of videosurveillance systems have been widely discussed during the last years [33],[27],[26],[6],[9], the problem has not been frequently addressed from the technological point of view. As recently noticed in [14], progressive advances in computing power and computer vision can help to achieve the right balance between privacy and security for videosurveillance. To the best of our knowledge, the first relevant attempt in this direction is due to the IBM Research Divi-

sion, as described in a technical report in 2003 [31]. Prior to its publication, only a few works had proposed technological solutions for privacy protection in the videosurveillance scenario (see [31], Sect. 3.3).

In [31], the authors propose the use of computer vision in order to understand the captured video and hide the sensitive data at different levels, according to the privileges defined for different users of the system. These privileges are to be defined considering the different classes of users of the system, which is dependent on each particular scenario. In an illustrative use case, three different classes of users are envisaged: anonymous (have access only to statistics collected by video analysis engines), privileged users (can watch video but with certain sensitive information hidden), and superusers, such as law enforcement officers (can watch the whole video without restrictions).

The privacy-preserving viewing of the video is ensured by means of a secure video console that re-renders the raw video, producing a new video stream where private data is hidden or obfuscated, whilst keeping the necessary information such that a human operator can evaluate the scene. Proper access control mechanisms guarantee that each user class has access only to the apropriate video stream. Thus, a security guard for instance could watch the video but with the faces of the recorded people erased or downsampled in order to make idenfication of the individuals impossible or at least very difficult. "Smart" video engines can detect different classes of objects (e.g. car plates), situations (e.g. accidents, fights, etc.) and individuals. In case the video engine is equipped with facial recognition capabilities, it is even possible to define different actions for the different individuals, provided the latter have been previously enrolled in the system.

The work described above gave rise to a significantly large number of papers, but mostly (or exclusively) dealing with implementations of video engines for detecting private data in the video stream and/or re-rendering raw video in a privacy-preserving manner. Some of these works will be briefly reviewed in Section 5.1.

3.3 Automated event control

Albeit automatic scene understading has been proposed as a tool to enable privacy-enhanced videosurveillance [31], up to now it has been raising more concern than relief, as it simplifies the collection and analysis of sensitive information on individuals.

Currently, a large number of tools for automating event control in videosurveillance, based on computer vision and image understanding, are being developed or already commercialized. Despite the large availability of tools of this kind, their massive deployment is still being hindered, not by the aforementioned privacy concerns, but rather by interoperability issues. In this context, a consortium originally formed by Axis Communications, Bosch Security Systems and Sony Corporation has began to promote the Open Network Video Interface Forum (ONVIF) standard [10]. More than 60 partners, including major actors in the videosurveillance industry, have already joined the initiative. ONVIF is aimed at developing an open standard for the communication between network video clients and video transmitter devices, giving a solution to the interoperability problem. The current ONVIF specification covers aspects such as device management, audio and video streaming, event management and video analytics, with all the interfaces described as web services by means of well known standards like XML, SOAP, and the Web Service Description Language (WSDL).

ONVIF defines two main architectural elements: the Network Video Transmitter (NVT), and the Network Video Client (NVC). The NVT is a device that sends video over an IP network to an NVC, so it plays the role of "service provider." On the other hand, the NVC is a controller device that communicates with an NVT, thus playing the role of "service requester." ONVIF gives support to JPEG, MPEG-4, H.264, G.711, G.726 and AAC codecs for video and audio streaming. The architecture of a video analytics application is composed of two main modules:

1. A video analytics engine which receives a video stream and produces a Scene Description, i.e. an abstract representation of the observed scene in terms of the objects present and their behavior.

2. A rule engine which contains the set of rules that govern the allowed actions in the observed scene (for instance, a certain virtual perimeter must not be crossed by pedestrians), the allowed intra-object relations (e.g. a person who lifts his/her bagage in the airport), and the allowed object behaviors (such as a maximum speed limit). The comparison between the Scene Description and the rules produces an event.

Clearly, ONVIF can have a great impact in the development of incoming privacy-preserving videosurveillance systems, as it opens the door to standard and interoperable management of sensitive events and privacy rules.

4. RIGHTS MANAGEMENT: A GLOBAL SOLUTION FOR VIDEOSURVEILLANCE

DRM has traditionally been used for protecting the rights of content creators. As pointed out in [23], due to the duality between rights management and privacy protection, DRM can also be applied, with some considerations, to privacy protection. The solution we propose goes even further. We have already anticipated that in a videosurveillance system, automated surveillance and privacy are closely related; thus, both aspects should not be considered independently, but jointly in a complete approach like the one we present. With a slight redefinition of some concepts, DRM can cope with all the aspects of videosurveillance, ranging from automatic event reporting to security of transmissions and storage, hierarchical access control for authorized users, and different levels of privacy protection for the surveilled users. At the same time, our proposal circumvents the problems presented by other individual privacy or automated surveillance systems (cf. Section 3) in terms of scalability, trust support or flexibility, always complying with the current regulations.

In the following we will describe the proposed architecture, depicted in Figure 2, of a videosurveillance system using DRM for providing both privacy and automated surveillance.

4.1 Object-users and Subject-Users

A videosurveillance system conceptually divides its users in two categories:

- **Subject-users**: We will call subject-users to those agents that have access to data generated by the system, and can perform actions on these data. In order

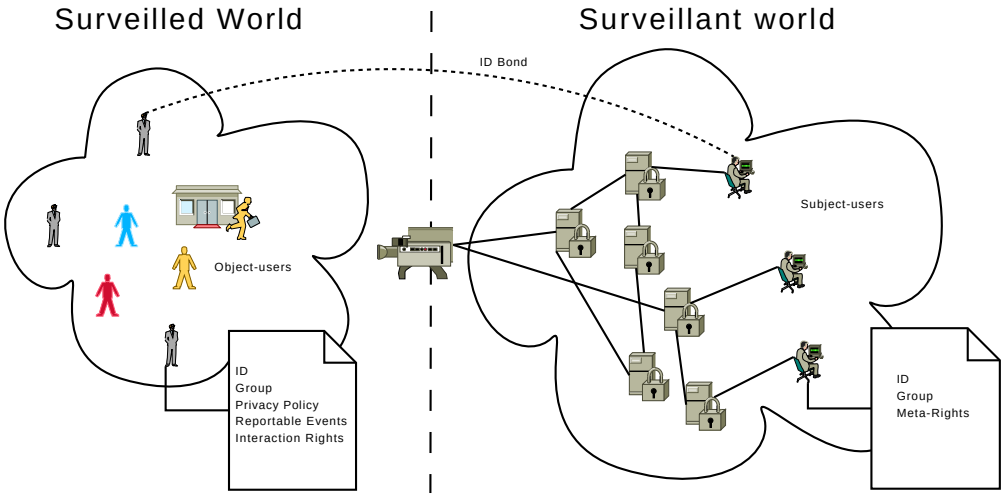

Figure 2: Architecture of a videosurveillance system using DRM

to access the system, they must authenticate themselves. In a DRM system, these users are represented as content consumers and adapters, that have certain access rights to the contents in the system, specified by the licenses of those contents. The information that these users produce is limited to event reports generated by the actions they perform, and contents adapted from existing contents in the system (when they have the right to produce them), but these users cannot generate new contents by themselves.

- **Object-users**: We will call object-users to those watched items (objects, people or regions), whose actions are surveilled and generate data. The data generation is initiated automatically when one of these users is recognized by the system in the contents generated by a camera; there is no authentication for these users besides the automatic recognition performed by the system; thus, there must be at least one object-user or group of object-users to which the detected but unrecognized users are mapped; it is also possible to automatically enroll every unrecognized user with a generic *unidentified* profile, in order to allow for relative ID recognition.

  Every object-user has an associated *virtual* content, that represents the object-user as an element of the DRM system with which other object-users may interact. It contains no resources, but only license and reporting information, (the *interaction rights* defined in Section 4.2). In the DRM system, these users are identified with content creators, and their associated *virtual* content is considered just as one ordinary content.

Informally, subject-users are those who can operate some part of the system, including employees of the enterprise, security guards, law authorities, etc. On the other hand, object-users are those in front of the cameras, that are being surveilled. Both categories of users must be uniquely identifiable by the system, and even when there is no a priori direct relation between them, the system may keep unique bonds between a determined subject-user and an object-user, that link both users under a unique system identity. For example, a security guard can be a subject-user, for he can have access to data in the system, but when appearing in front of the camera, the guard becomes also an object-user, although both users are bond to the same identity. These links allow for the assignment of access rights of a subject-user to the data generated by the linked object-user; thus, it solves the problem of automatically providing a recognized object-user access to his/her own data without the need of an external process that determines the relationship between the data and the authorized subject-user; this would effectively implement the data access right required by Directive 95/46/EC. Nevertheless, these links constitute only identity links, and access rights can be defined elsewhere, as will be shown later on. This is another example of the need of taking into account the dependency between privacy and automatic surveillance.

4.2 Roles and privileges

As we are dealing with a global solution that covers all the aspects of videosurveillance through a DRM system, we must take into account two complementary questions:

- **Privacy**: When taking into account the privacy concerns related to videosurveillance, it is customary to protect the information of the object-users that are being recorded, and allow access to this information only to authorized subject-users. This implementation through DRM would improve on the typical access control lists commonly proposed for privacy protection, providing a more flexible access system and different access levels. Regarding privacy, it is also desirable to implement different privacy profiles, such that each user has relative freedom to choose how, when and against whom his/her data must be protected, and whether or not the object-user wants to be informed whenever some subject-user accesses his/her data. On the other hand, subject-users generate only reports on their activity on data previously created by object-users; thus, the privacy of subject-users is taken into account by granting them privileges in order not to report certain actions; for example, a representative of

the law authorities may have visioning access to some recorded scenes, and this event should not originate reports to the involved object-users.

- **Event control**: An automated surveillance system must be able to detect certain events and inform to the appropriate agent(s) of the system about their occurrence. In this way, the system must define which actions of object-users must be subject to event control, and which subject-users must be informed about those actions; this should be either a generic policy, affecting all object-users, or a specific policy, affecting a limited group of users (or only one user). E.g., an enterprise may want that several members of the "security" group be informed about determined events happening between object-users; this would define a policy affecting all the involved object-users.

This two-fold approach can be materialized through the application of DRM with the following mapping:

- **Object-users**: For every object-user, the following elements will be defined:
  - **Privacy policy**: The object-user, as a content generator of the DRM system, must have a data creation policy (a template for content generation and for the associated licenses), that defines:
    - The access rights that each group of subject-users has over each piece of information generated by the object-user in question.
    - A selection of tools used to secure the generated contents. This selection will represent the way the data will be protected, and therefore, the privacy level assigned to the user; they can comprise, but are not limited to: distortion, masquerading, substitution, encryption or even total elimination; this last choice would represent the level of total privacy, for which unauthorized users would not have access to the data and also to knowing whether some data is present or not. More about total privacy is discussed in Section 5.
  - **Reportable events**: A list of actions (record, play, copy, encapsulate,...) on the data generated by the object-user that must generate event reports, and to which subject-users these reports are sent.
  - **Interaction Rights**: The surveillance system may be able to detect interactions between object-users (i.e., a person entering a restricted access zone, a person leaving a case on the floor, a fight,...). Each object-user or group of object-users may have a list of allowed/forbidden interactions, that will define which interactions must be recorded as an event report, which the level of those reports is, and to which subject-user(s) those reports will be sent. This rights are defined in the *virtual* content associated to each object-user.
  - **Group**: Every object-user should belong to one of the defined groups, as Interaction Rights are more efficiently described when specifying a group and an object than when specifying pairs of objects. Additionally, groups of object-users may have default privacy policies. Grouping is a critical aspect to allow for scalability of the system.
- **Subject-users**: Subject-users have their privacy taken into account through restrictions on the reported events that their actions generate, granting them privileges in order not to report certain actions.

  As well as object-users, subject-users can also be grouped, as there may be some generic classes of users (like representatives of the law, security professionals, regular employees,...), with default profiles and different privileges. Additionally, subject-users may also have meta-rights over virtual contents, being able to issue or revoke the access rights of some object-user or group of object-users to certain virtual contents.

4.3 Event Management

As highlighted in the previous sections, there are two types of events: those produced by the interaction between object-users, and those produced by the actions performed by subject-users on system data. The former will be called *surveilled events*, while the latter will be called *privacy events*. The reports that these events generate will always have one or several subject-users as recipients. Nevertheless, all of them can be handled in a unified way by a DRM system supporting event reporting or metering functions.

Events are actions that the system can identify. *Privacy events* are defined by typical access rights of subject-users to ordinary contents of the system. This is directly interpreted as traditionally done in a DRM system. On the other hand, *surveilled events* are defined by access rights of object-users to *virtual* contents. Thus, the surveillance system must identify *objects* in the surveilled scene, map them to their corresponding object-users and virtual contents in the DRM system, and map the interaction among objects to actions performed by object-users on virtual contents; then, the virtual contents will indicate which of these actions must be reported as events.

4.4 Surveillance Ontology

As shown in the previous sections, the framework we propose can cope with both sides of a videosurveillance system: the world behind the cameras, and the world in front of them. Thus, the ontologies typically associated to DRM systems are not enough to describe all the agents and relationships that are present in our framework, such as the interaction rights. Therefore, an extended ontology is needed to cope with the concepts and the bonds that are present in our framework.

5. IMPLEMENTATION GUIDELINES: A HIGH LEVEL PERSPECTIVE

5.1 Prior works

Most works on videosurveillance privacy are very recent, and most of them are focused on particular modules of the global privacy management architecture, especially on: 1) engines for automated detection of objects of interest; and 2) methods for protection (including blurring and encryption) of the private data.

As for the first module, the objects of interest are usually chosen to be persons and moving objects such as vehi-

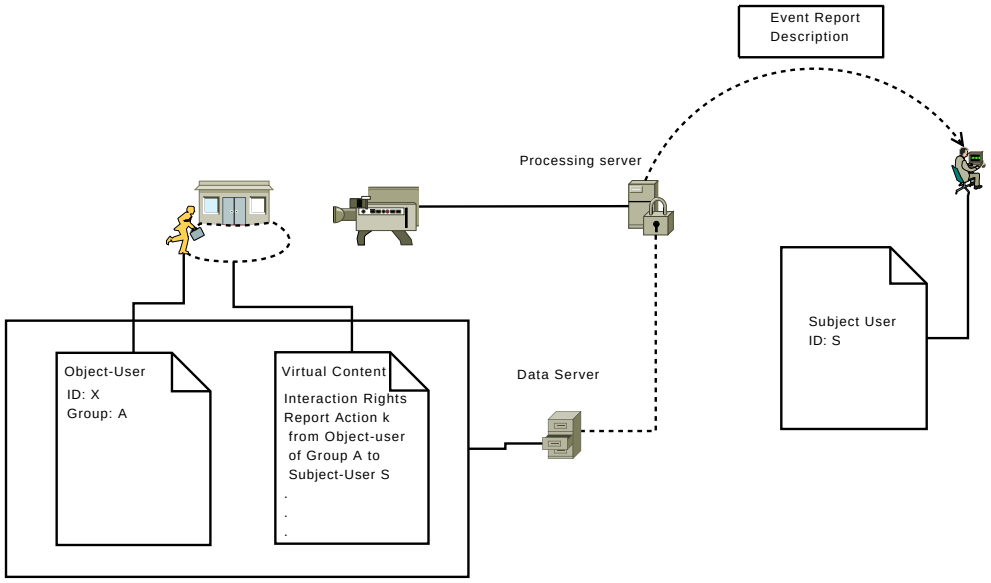

Figure 3: Architecture of event reports using DRM

cles. The main reason is simplicity, since the considered scenarios usually have static videosurveillance cameras: hence, moving object detection can be accomplished by means of simple background segmentation techniques [15], [18], [38], [32]. More sophisticated approaches for object detection are based on trained classifiers, such as [25], [28], [16]. These approaches allow for better control and decision making, but at the cost of increased computational complexity. Nevertheless, a bunch of efficient detectors are described in the literature and are successfully used in real scenarios, like the method by Gavrila and Philomin [21], the well-known Viola-Jones method [34], or the HOG method [17].

Regarding the protection of private data, we can find works such as [29], [16], [32], [38], where a variety of non-invertible transforms are applied to the sensitive data: obscuring, pixelization, blurring, silohuetting, face masking, etc. In other works, the application of standard encryption methods (e.g. AES, permutation-based encryption) is proposed in order to ensure recoverability of the private data if the viewer has access to the appropriate key. Some works simply apply encryption to the raw bits comprising the detected objects of interest [15], but others resort to layered encryption techniques combined with perceptual coding, in order to provide different privacy levels in a natural manner. Two examples can be found in [18], [24] for Motion JPEG-2000 video. Other works on layered encryption can be found for H.264 [30], DCT-based [19], Hierarchical MPEG [22], and scalable codecs in general [20]. However, the four latter works are not directly applicable to the videosurveillance scenario, as no objects of interest are considered (the whole image is encrypted).

In general, the usability of the proposed encryption algorithms is strongly dependent on the video codec used in the videosurveillance system (especially in layered encryption techniques). In [13], permutation-based image encryption is proposed in order to achieve codec independence. Furthermore, this encryption method allows for a certain transcoding of the video without completely destroying the encrypted information if afterwards recovered using the proper secret key. The system performance has been evaluated with several video codecs, including MPEG-2 and H.264.

In view of the existing work, it appears that no global framework for a practical implementation of privacy management solutions exists for the videosurveillance scenario. Our proposal in this direction is introduced below.

5.2 A generic, standards-compliant implementation proposal

At a high-level, our implementation proposal can be viewed as a generic video analysis system coupled with a DRM system. It is basically a combination of MPEG-4, MPEG-21 and ONVIF standards, in such a way that all aspects of the proposed architecture and functionalities are satisfactorily covered:

- MPEG-4 covers aspects related to object management and cryptographic protection of the sensitive data (IPMP);
- MPEG-21 provides standard means for defining licenses on data use and a language for expressing rights (REL), as well as the format for event reports and their requests.
- ONVIF provides event management and transmission capabilities.

The different components of our proposal are described below. It must not be understood in any way as a restrictive implementation, but just as an illustrative implementation which is generic enough and covers all aspects of the proposed architecture using standard technology.

5.2.1 Video segmentation, encryption and encoding using MPEG-4

Unlike MPEG-2 and MPEG-1, MPEG-4 strongly relies on the concept of object. This is one of the characteristics that makes it specially amenable to the implementation of our proposal. Whereas an MPEG-2 program is typically formed by two audiovisual elements (one full-screen video stream,

and one audio stream), MPEG-4 content may be built of an arbitrary number of audiovisual elements, called *objects*, that belong to a wide range of defined object types, such as rectangular video, video with shape, synthetic face or body, speech, synthetic audio, text, or graphics. The basics of the object encoding mechanisms available in MPEG-4 are briefly explained in the following.

As illustrated in Figure 4, the access to MPEG-4 contents starts with an Initial Object Descriptor (IOD). This IOD points to at least two basic streams: a scene description (BInary Format For Scenes, BIFS), and an Object Descriptor (OD) stream. The OD is a kind of container aggregating all the useful information about the corresponding object. An OD can contain a URL pointing to a media stream, or a series of subdescriptors. These subdescriptors contain pointers to individual Elementary Streams (ESs), semantic information about an object, and pointers to contents access management information (Intellectual Property Management and Protection, IPMP). IPMP descriptors and IPMP streams specify a means for decrypting ciphered ESs, or for checking authorization or entitlement information. It is worth pointing out that a single visual or audio object can be coded into one or more ESs.

OD streams are usually associated with a scene description (BIFS) stream; the scene description conveys the spatiotemporal layout of the media objects in the scene, i.e., it indicates how to assemble the various media streams described within the OD stream. ODs and BIFS are associated through another OD. Based on BIFS, a visual scene in MPEG-4 is described as a composition of Video Objects (VOs) characterized by their shape (not just rectangular, but arbitrary shapes can be considered), motion, and texture. Each VO can consist of one or more layers (VOL) which can be used to enhance the temporal or spatial resolution of a VO. An instance of a VOL at a given time instant is called a Video Object Plane (VOP).

Due to its hierarchical structure, its modular nature and the diversity of its encoding tools, MPEG-4 indeed provides many degrees of freedom for producing a video sequence. This high number of possibilities can be effectively exploited in our framework for videosurveillance.

As explained before, we are interested in hiding certain parts of the video stream, depending on the rights of the viewer and the licenses associated to the objects in the image.[1] MPEG-4 allows to define a bottom-to-top video structure, where it is possible to separately encode and protect the objects of interest, such that they can be a posteriori selectively extracted and reproduced, according to the relevant rights and licenses. An illustrative example of these capabilities is the following.

1. In the videosurveillance scenario, the first object of interest to be defined is the background image, usually static, being captured by the camera. For still cameras, this background image can be easily defined. For motorized remotely controlled cameras, a reference background image can be also defined for each possible positioning vector (including azimuth and elevation parameters) and configuration parameters (focus, shutter time, resolution, etc.)

2. The background image plays the role of canvas over which the moving objects (e.g. persons, vehicles) are superimposed in order to compose the complete scene consisting of a set of layers (VOL).

3. The viewer (subject-user) can check in the IPMP descriptors whether he/she has the rights to visualize the objects in the scene (in the clear). If that is the case, the corresponding VOs and VOLs will be conveniently decoded, decrypted and placed in the scene thanks to the BIFS capabilities. On the contrary, for the subject-user lacking the necessary rights, the system would render a processed version of the protected objects over the background image. This processing may encompass partial decryption, blurring, transparency (even total erasure), substitution by a synthetic face/body, etc., according to the degree of privacy required and/or the rights of the subject-user. An example of the output presented to a subject-user is shown in Figure 5.

This conditional access problem can be solved by means of traditional DRM tools applied on the media streams needed for the composition of the final video. The application of DRM can be performed at different points of the MPEG-4 structure.

- A simple approach would consist of encrypting the ESs corresponding to the media objects representing the object-users of interest, and sending them by default in a stream available to all the subject-users. One obvious drawback is the significant communications overhead (the encrypted data is useless for unauthorized viewers), and the possible privacy leaks because of the mere awareness that certain information is being hidden.

- A more sophisticated approach would be the use of encrypted URLs in the ODs pointing to the DRM-protected streams, that could be even stored in a different video server. This way, a twofold objective would be achieved: 1) unauthorized users would not download data they would not be able to process, and 2) privacy would be improved to a larger extent, as unauthorized users will know that some encrypted URLs are present, but they will not be able to know exactly how many there are. Furthermore, strategies could be designed where a media object were partitioned into multiple ESs, each of them DRM-protected and stored in a different server; doing so, not even hacking a server would be enough for accessing the protected data, as the set of all the servers containing data of a multimedia object should be cracked. Therefore, the privacy and security levels provided by this approach would be significantly higher than for the privacy-enhancing videosurveillance systems proposed so far.

5.2.2 *Licensing, rights and event management implementation via MPEG-21 and ONVIF*

MPEG-21 provides a rich and extensible language for defining user rights and allowed uses of digital contents, specified as a Rights Expression Language (REL). These rights are introduced in licenses that, when associated to a Digital

[1]Note that the MPEG-4 objects of interest in our scenario are content media generated by object-users (i.e., multimedia information containing the image/voice of the object-user), so the fact that a given subject-user has the appropriate rights to access the data will not just depend on the subject-user group and the corresponding rights, but it will also depend on the object-user privacy policy.

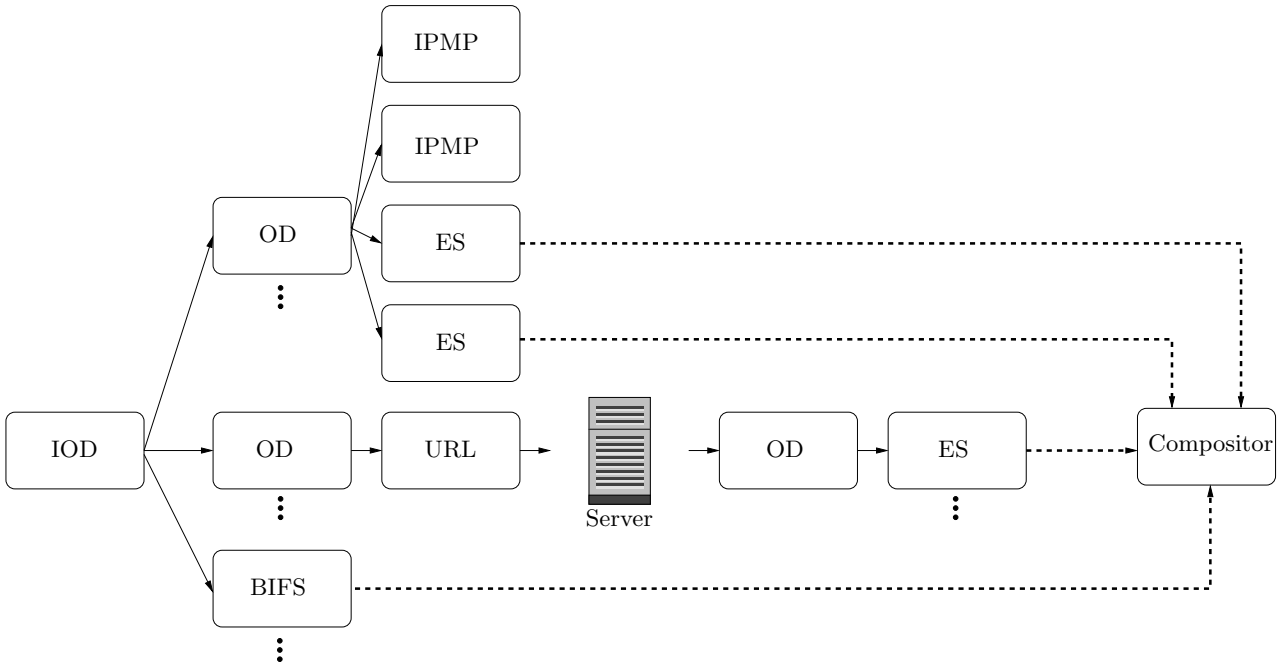

Figure 4: Schematic description of MPEG-4 contents based on objects.

Item (DI), specify the actions that determined users can perform on the resource linked to the DI. The rights defined in MPEG-21 REL can be directly mapped to the actions (play, record, copy,...) performed by subject-users on the contents generated by the videosurveillance system; nevertheless, the REL must be extended in order to include the actions that object-users perform on virtual contents, e.g., enter, leave, take,... (cf. Section 4). MPEG-21 REL allows to easily define these extensions.

On the other hand, the ONVIF standard provides a set of security operations for configuring NVTs including, among others, setting access security policies, and handling user credentials and settings. The security model in ONVIF ([8]) defines a standard set of SOAP extensions, and includes a Token Profile based on REL ([7]), considering four different user levels: Administrator, Operator, Media user, and Anonymous. The access security policy of each of these groups can be defined through REL. Thus, both standards can complement each other within an implementation of our framework: while MPEG-21 REL provides the standard language for defining user rights and access privileges, ONVIF would provide standardized protocols for the protection of these rights and privileges.

Event handling in ONVIF is based on the OASIS WS-BaseNotification and WS-Topics specifications [2], that define event handling principles, basic formats and communication patterns; however, the standard does not require particular notification topics, and it defines a set of basic notification topics that an NVT is recommended to support [10]. Likewise, MPEG-21 Event Report (ER) defines the format of the Event Report Request (ERR) and Event Report (ER) messages. It must be noted that both standards do not collide also in the specification of event handling, as MPEG-21 ER deals with report formats, while ONVIF defines the layer for communicating generic reports (Real-Time and non Real-Time). Thus, turning again to our framework, a natural implementation of event reporting would consist in using MPEG-21 formats for ER and ERR, embedding the latter into the corresponding DIs, and encapsulating the former inside ONVIF messages, in order to be simultaneously compliant with both standards.

6. USE CASE DESCRIPTION

This section describes by example the application of the videosurveillance model presented in Sect. 4 to a real scenario. We focus on a videosurveillance system for airports because such scenario encompasses a wide range of users and situations that illustrate well the capabilities of the proposed approach.

6.1 Users

According to Sect. 4.1, we distinguish between object-users and subject-users. Object-users are those individuals/entities in front of the cameras, and some examples are

1. public security forces (police and similar),
2. private security agents,
3. airport assistants,
4. airlines desk staff, airlines on-board crew,
5. duty free, bars and restaurants or cleaning staff,
6. airport users (e.g. passengers),
7. baggage items and trolleys,
8. security control trays,
9. restricted areas.

(a) Video frame in the clear

(b) Privacy-protected video frame

Figure 5: Example of privacy-preserving videosurveillance by means of person detection and pixelization.

On the other hand, subject-users are those who operate the videosurveillance system. In this case, it would correspond to user types 1 and 2. Only user types 1-6 may define how the information generated by them is processed, while user types 7-9 will be assigned a generic or specific policy.

The classification of a given object-user in one of the aforementioned categories requires a previous enrollment and/or system training. In the case of items or areas (user types 7-9), the videosurveillance system needs a training phase to understand what kind of object it has to look for, and sometimes intervention of a subject-user (to define the perimeter of a restricted area, for instance). For many human object-users (types 1-5), enrollment is already usually required in current airport security systems. It is not necessarily the case for passengers and other airport users, who can anonymously move around many areas inside the airport. Thus, they would belong in general to the class of *unidentified* object-users. Nevertheless, the fact that a passenger does not follow a typical enrollment process does not imply that he/she can not be tracked all over the airport. The videosurveillance system could have a database containing information about the observed passengers in the last, say, 24 hours. This way, a high level of security can be achieved whilst complying with the proportionality principle.

Notice that human object-users of types 1-5 are usually required to bear an identity card, which can be equipped with an RFID device, and hence used for verifying unobtrusively their identity against a biometric recognition system integrated in the videosurveillance network.

6.2 Privacy policies

According to Sect. 4.2, object-users can define through their privacy policies how their information is managed by the system and accessed by others. Some examples are given here:

1. A shop assistant in a duty free store can request to make his/her image unidentifiable to private security agents operating the videosurveillance system while he/she is at work inside the area of the store facilities.

2. A police officer can request his/her image to be completely removed for any subject-user, unless it is another police officer. In addition, he wants his private data to be protected with RSA encryption.

3. Default rules can be applied to unidentified object-users, such as passengers and other airport users not enrolled in the security system.

These privacy policies will be eventually reflected on the access rights that are granted to subject-users, depending on their specific function at the airport. For example, subject-users of types 1-2 have access to most of the information that is processed/recorded by the videosurveillance system. On the other hand, subject-users corresponding to object-user types 3-6 (or even user types 1-2 not in charge of operating the videosurveillance system) should not have access to these data.

6.3 Interaction rights and automated event control

According to Sect. 4.2, object-users have predefined rights that control the allowed interactions between them. The automated videosurveillance system can detect whether these interaction rights are being violated in order to appropriately generate event reports. Some examples are:

- A piece of baggage (type 7 user) remains unattended for a long time. The system tags this item as "unattended baggage" and sends the correspoding report to a security officer.

- A type 6 user picks up a piece of baggage that had been tagged as "unattended baggage". This could indicate a theft, so the appropriate report is generated and sent again to the security officer.

- A type 6 user enters a restricted area (type 9 user). The system checks whether the former has the right to access this area; if not, the action is notified via an event report.

7. CONCLUSIONS AND FINAL CONSIDERATIONS

Individuals demand for technical ways of improving their personal security that do not hinder their right to privacy,

which is granted by the European legislation in force. As far as videosurveillance is concerned, the DRM-based architecture that has been proposed in this paper comes to provide a good balance between security and individuals' privacy. In fact, as presented, DRM can solve the critical issues of a current videosurveillance system, considering their twofold nature: covering and standardizing the automation of the surveillance activity, while putting in the hands of the users the appropriate technical means to control the access to their private information. Thus, the way DRM is used in our framework can lead to an increased acceptance of videosurveillance, as its target is the protection of the final user.

Note that this paper has dealt only with conceptual elements of a videosurveillance network, without taking into account how they can be mapped to physical devices. The most straightforward solution is to include in the network dedicated processing nodes, which can provide the necessary analysis and automated decision functionalities. On the other hand, the computation power of IP "smart" cameras is rapidly increasing, so it is foreseeable that videosurveillance cameras will soon become autonomous devices capable of performing complex video processing operations. This will ensure that many of the functionalities envisaged by the presented architecture will be directly realizable in the cameras. Additionally, the more powerful the analysis engines run by the system, the more granular and reliable it will be.

Finally, it has been shown that the presented architecture can be implemented through the use and adaptation of current standards, like ONVIF, MPEG-21 and MPEG-4. This constitutes a clear advantage, in the sense that a standards-compliant solution provides more generality, transparency and availability.

8. REFERENCES

[1] The extensible rights markup language. http://www.xrml.org.

[2] Oasis standards and other approved work.

[3] P3P: The Platform for Privacy Preferences. http://www.w3.org/P3P/.

[4] Directive 95/46/EC of the European Parliament and of the Council. Official Journal L 281, 23/11/1995 P. 0031 - 0050, October 1995.

[5] Organic law 15/1999, on the protection of personal data, December 1999.

[6] Privacy review: video surveillance programs in Peterborough. Information and Privacy Commissioner Office, Ontario, December 2004. Report.

[7] Web Services Security Rights Expression Language (REL) Token Profile 1.1, February 2006. OASIS Standard.

[8] Web services security: Soap message security 1.1 (ws-security 2004), February 2006. OASIS Standard.

[9] Under the watchful eye: the proliferation of video surveillance systems in California. Americal Civil Liberties Union (ACLU), August 2007. Report.

[10] ONVIF core specification ver 1.0, November 2008.

[11] S. D. P. Agency. Instruction 1/2006, on processing personal data for surveillance purposes through camera or video-camera systems, November 2006.

[12] A. Becker, A. Arnab, and M. Serra. Assessing privacy criteria for drm using eu privacy legislation. In *Proceedings of the 8th ACM workshop on Digital Rights Management*, pages 77–86, Alexandria, Virginia, USA, October 2008.

[13] P. Carrillo, H. Kalva, and S. Magliveras. Compression independent object encryption for ensuring privacy in video surveillance. In *IEEE International Conference on Multimedia and Expo*, 2008.

[14] A. Cavallaro. Privacy in video surveillance [in the spotlight]. *IEEE Signal Processing Magazine*, 24(2), March 2007.

[15] A. Chattopadhyay and T. Boult. Privacycam: a privacy preserving camera using uCLinux on the Blackfin DSP. In *IEEE Conference on computer vision and pattern recognition, CVPR'07*, Minneapolis, MN, USA, 17-22 June 2007.

[16] D. Chen, Y. Chang, R. Yan, and J. Yang. Tools for protecting the privacy of specific individuals in video. *EURASIP Journal on Advances in Signal Processing*, 2007:1–9, 2007.

[17] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. volume 1, pages 886–893, Los Alamitos, CA, USA, 2005.

[18] F. Dufaux, M. Ouaret, Y. Abdeljaoued, A. Navarro, F. Vergnenègre, and T. Ebrahimi. Privacy enabling technology for video surveillance. In S. S. Agaian and S. A. Jassim, editors, *Mobile Multimedia/Image Processing for Military and Security Applications*, volume 6250. SPIE, 2006.

[19] M. M. Fisch, H. Stögner, and A. Uhl. Layered encryption techniques for DCT-coded visual data. In *European Signal Processing Conference on Signal Processing, EUSIPCO'04*, 2004.

[20] C. Fonteneau, J. Motscha, M. Babela, and O. Déforges. A hierarchical selective encryption technique in a scalable image codec. In *International Conference in Communications*, Bucharest, Romania, 2008.

[21] D. Gavrila and V. Philomin. Real-time object detection for "smart" vehicles. In *Seventh IEEE International Conference on Computer Vision*, volume 1, pages 87–93, 1999.

[22] H. Hofbauer, T. Stütz, and A. Uhl. Selective encryption for hierarchical MPEG. In S. B. . Heidelberg, editor, *Communications and Multimedia Security*, volume 4237/2006 of *Lecture Notes in Computer Science*, pages 151–160, 2006.

[23] S. Kenny and L. Korba. Applying digital rights management systems to privacy rights management. *Journal of Computers and Security*, November, 2002.

[24] K. Martin and K. N. Plataniotis. Privacy protected surveillance using secure visual object coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 18(8):1152–1162, 2008.

[25] I. Martínez-Ponte, X. Desurmont, J. Meessen, and J.-F. cois Delaigle. Robust human face hiding ensuring privacy. In *International Workshop on Image Analysis for Multimedia Interactive Services, WIAMIS'05*, Montreux, Switzerland, April 2005.

[26] M. McCahill. *The Surveillance Web: The Rise of Visual Surveillance in an English City*. Willan Publishing (UK), 2002.

[27] M. McCahill and C. Norris. Cctv in London. Report deliverable of UrbanEye project, 2002.

[28] Q. Meibing, C. Xiaorui, J. Jianguo, and Z. Shu. Face protection of h.264 video based on detecting and tracking. In *8th International Conference on Electronic Measurement and Instruments, ICEMI'07*, volume 2, pages 172–177, 16-18 July 2007.

[29] E. M. Newton, L. Sweeney, and B. Malin. Preserving privacy by de-identifying face images. *IEEE Transactions on Knowledge and Data Engineering*, 17(2):232–243, February 2005.

[30] S.-W. Park and S.-U. Shin. Efficient selective encryption scheme for the H.264/Scalable Video Coding(SVC). In *Fourth International Conference on Networked Computing and Advanced Information Management*, volume 1, pages 371–376, Gyeongju, Korea, September 2-4 2008.

[31] A. Senior, S. Pankanti, A. Hampapur, L. Brown, Y.-L. Tian, and A. Ekin. Blinkering surveillance: enabling video privacy through computer vision. Technical report, IBM Research Division, August 2003.

[32] S. Tansuriyavong and S. ichi Hanaki. Privacy protection by concealing persons in circumstantial video image. In *ACM Workshop on Perceptive user interfaces*, 2001.

[33] N. Taylo. State surveillance and the right to privacy. *Surveillance and Society*, 1(1):66–85, 2002.

[34] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 511–518, 2001.

[35] X. Wang, T. DeMartini, B. Bragg, M. Paravasivam, and C. Barlas. The MPEG-21 Rights Expression Language and Rights Data Dicitionary. *IEEE Transactions on Multimedia*, 7(3):408–417, June 2005.

[36] D. J. Weitzner, H. Abelson, T. Berners-Lee, C. Hanson, J. A. Hendler, L. Kagal, D. L. McGuinness, G. J. Sussman, and K. K. Waterman. Transparent accountable data mining: New strategies for privacy protection. In *AAAI Spring Symposium on The Semantic Web meets eGovernment*, 2006.

[37] A. Westin. *Privacy and freedom*. The Bodley Head Ltd., 1970.

[38] X. Yu, K. Chinomi, T. Koshimizu, N. Nitta, Y. Ito, and N. Babaguchi. Privacy protecting visual processing for secure video surveillance. In *IEEE International Conference on Image Processing, ICIP'08*, pages 1672–1675, San Diego, CA, USA, 12-15 October 2008.