Enforcing Privacy through Usage-Controlled Video Surveillance[*]

Pascal Birnstill
Fraunhofer IOSB, Karlsruhe, Germany

Alexander Pretschner
Technische Universität München, Germany

Abstract

Increasing capabilities of intelligent video surveillance systems require the enforcement of privacy-related requirements. Data usage control technologies offer appropriate solutions in this problem domain. We first present specific requirements for a privacy enforcement infrastructure for modern surveillance systems that we align with a generic architecture and a privacy-aware workflow template for operating such systems. To ensure the compliance of a surveillance system's operation with such a workflow, we then derive respective usage control requirements. We show that the conceptual framework of usage control provides suitable instruments for specifying these requirements and for implementing the corresponding enforcement mechanisms. Our architecture has been implemented prototypically.

1. Introduction

Intelligent video surveillance is an active field of research, predominantly in the domains of image exploitation and situation assessment algorithms. The availability of privacy-invasive system functionality such as real-time object tracking and automatic extraction of biometric features is becoming reality. Not surprisingly, video surveillance generates an increasing interest in information security and privacy.

A categorical argument against video surveillance targets the panoptic effect of such systems, which arguably is in conflict with the fundamental right to free development of the individual. When faced with surveillance cameras, we cannot know whether or not we are currently observed. However, the mere possibility of being observed tends to change the way we behave, which usually is considered an undesired phenomenon in free societies and therefore addressed by legislation. The principle of proportionality as laid down in articles 8(2) and 52(1) of the Charter of Fundamental Rights of the European Union demands a careful weighing of the purpose of a surveillance measure, *i.e.*, the legally protected interest to be defended, against the legitimate interests of people affected by the surveillance measure. However, we do observe that video surveillance is spreading rapidly, even though the proportionality of privacy invasion and utility may not always be justified.

In addition, even if we consider video surveillance to be lawful in particular cases, this raises the question of how and to which extent privacy of the people concerned can be preserved as far as possible, without interfering with the intended lawful purpose of a given surveillance measure: How can we design privacy-preserving mechanisms that do not render surveillance technology useless?

In this paper, we investigate how to design and implement technology for enhancing privacy in surveillance systems by weighing lawful utility against privacy. This seems to be particularly relevant given that modern surveillance technology works at the level of objects rather than video streams: video streams are fused into various objects including attributes such as IDs by face recognition, location, change of location, any activities. This technology also enables automated tracking of objects across cameras, thus turning the above privacy concerns even more convincing.

We tackle the following **problem**. How can we design privacy mechanisms for video surveillance systems that (1) work at the level of object streams rather than video streams and that (2) do not render these systems useless by over-emphasizing privacy over utility? Our **solution** is a generic camera surveillance architecture that enforces privacy requirements with data usage control technology. In terms of our **contribution**, the analysis of related work in §2 reveals that: (1) Some existing privacy mechanisms for camera surveillance systems work at the level of video streams rather than at the level of object streams into which video streams are fused. (2) Other mechanisms deactivate surveillance by default and only activate it when explicitly triggered, therefore making it impossible to track, for instance, suspicious luggage being dropped. (3) A third class of existing privacy mechanisms is inherently bound to an observation purpose which, specifically so in publicly deployed surveillance systems, is hard to render operational. Our contribution is privacy enhancing technology for surveillance cameras that

[*] Work done as part of the Fraunhofer-Gesellschaft internal program Attract 692166 "Distributed Data Usage Control."

(1) works on object streams, (2) is always switched on yet privacy-friendly as long as no alarm is automatically triggered, and (3) is not directly bound to a purpose that would be hard to capture, but rather relies on the notion of distinct operational modes. As a proof of concept, we implemented usage control enforcement for several components of the video surveillance testbed *Network Enabled Surveillance and Tracking (NEST)* [6]. A demo video is available at http://www22.in.tum.de/fileadmin/demos/uc/Demo4-UC4NEST-Demo-v3.mp4.

2. Related Work

Privacy for Surveillance Cameras. We concentrate on work on privacy enforcement in video surveillance systems because our work is orthogonal to computer vision techniques for enhancing privacy.

Fidaleo et al. propose a privacy-enhanced surveillance architecture in which a so-called privacy buffer detects and removes identifiable information, *e.g.*, persons' faces, from input data [2]. The operator is granted interactive control over certain system functions. This does not seem to be situation-dependent. Weighing the appropriateness of a surveillance measure against its intrusiveness, a system which is most of the time as little intrusive as possible is considered "better" than a system that persistently sticks to the same trade-off between privacy and utility. Aiming at reducing a surveillance system's privacy-invasiveness by default, we contribute mechanisms for restricting the usage of intrusive surveillance operations to the scope of alarms.

In [11] Senior et al. introduce a privacy-preserving video console for hiding sensitive details in video streams depending on authorization levels. This suggests that the privacy level of exposed video data should be adjusted exclusively to the authorization level of the observer, as opposed to the authorization level induced by the surveillance purpose.

Saini et al. [10] quantify the loss of privacy due to video surveillance recordings by decomposing embedded information into *what*, *when*, and *where* evidence. Such evidence may (1) be sensitive in case the person's identity is unveiled and (2) constitute context knowledge, which allows for drawing inferences about the identity. However, eliminating *when* and *where* evidence in addition to obfuscating personal features turns out to be hard in practice. Hence, for the time being we aim at exposing as little video data in the surveillance process as possible.

Wickramasuriya et al. enforce privacy policies for video re-rendering [14]. Surveillance is restricted to critical regions. Cameras are deactivated by default and activated by motion detectors if people enter such regions. Policies specify access rights to regions and privacy levels for individuals or groups. People are authenticated using RFID tags. When entering critical regions with an RFID tag granting access, one may also be granted a high privacy level, *i.e.*, be erased from visualized video data. This seems useful for surveilling people in constrained regions. However, even while staying in the observed area, people can transfer their identity to someone else by passing on their RFID tag. In contrast, our approach inhibits identity transfers, employing the system's tracking capabilities to persistently bind authenticated identities to captured objects. We cater to data protection of employees by enforcing policies specifying privacy-enhancing mechanisms on particular objects or object types, *e.g.*, hiding authenticated staff.

Mossgraber et al. have introduced the notion of task-based video surveillance [7], the benefits of which for privacy have been elaborated by Vagts and Bauer [12]. System functionality is decomposed into individual surveillance tasks, which are triggered on behalf of an authorized human user and are not supposed to exchange data among each other. Thus, aiming at data minimization, video data must only be acquired, processed and stored if required by an authorized task. This approach seems appropriate if the surveillance purpose does not require a significant extent of continuous image exploitation. Furthermore, in order to apply a task-oriented approach, either the principal purpose of the surveillance measure must decompose into distinct sub-purposes, or multiple purposes must be intended from the beginning. In practice, surveillance measures in public spaces are usually dedicated to a rather broadly conceived legal purpose requiring a broad spectrum of detective functionality. Exemplary purposes are fighting (some sort of) crime in public spaces or protecting civil security on an airport. A meaningful decomposition of this kind of purposes is not straightforward and does not directly seem to lead to increased privacy. Deploying a surveillance system for multiple distinct purposes is, for legal reasons, almost only conceivable for deployments in non-public environments, *e.g.*, in office buildings. In such scenarios, surveillance systems are typically utilized for monitoring critical areas, valuable objects, or on-demand tracking of persons, such as unknown visitors. However, surveillance systems in non-public environments constitute a small fraction of privacy invasions induced by video surveillance technologies. We favor a separation of system functionality into two operational modes (cf. §4) according to its detective or reactive nature and its intrusiveness over a separation into tasks.

Distributed Usage Control. Usage control (UC) generalizes access control to the time after initial access to data [8]. Requirements include rights and duties, *e.g.*, "data may not be forwarded," "data must be logged and deleted after thirty days," *etc.* UC requirements are specified in policies. In distributed settings, *e.g.*, forwarding a data item with an attached policy to another system, UC requirements can be enforced on the receiver's machine, too, requiring UC enforcement mechanisms at the receiving end [5].

Policies are usually specified via events. Because data usually comes in different representations—an image can be a pixmap, a file, or aggregated into the set of objects shown on the image—UC mechanisms have been augmented by data flow tracking technology [4, 9]. One can then specify policies not only for specific fixed representations of data, but also on *all* representations of that data. These representations are tracked by information flow detection technology. Policies then do not need to rely on events but can forbid specific representations to be created, also in a distributed setting [5]. To our knowledge, UC enforcement and data flow tracking technologies have not been applied to video surveillance systems yet.

3. A Generic Video Surveillance Architecture

We propose a holistic approach to privacy-aware operation of video surveillance systems. To this end, we introduce a generic privacy-aware workflow that stipulates how human users may operate such systems (cf. §4). In order to ensure compliance with this workflow, appropriate enforcement mechanisms must be established in the surveillance system's data processing chain. They must be able to control collection, exploitation, visualization and storage of data. We propose to use usage control technology (cf. §2).

With two operational modes, a default and an alarm mode, we restrict intrusive surveillance operations (*e.g.*, recording video, analyzing object tracks, biometric feature detection, *etc.*) to the scope of incident handling. We also improve the selectivity of video surveillance as we only allow these operations to collect data related to the person who provoked the alarm. Keeping track of the particular pieces of data concerning a given alarm, we ensure selective deletion (compliance with deadlines, false alarms) and encryption of sensitive data (preservation of evidence).

Orthogonal to the operational modes, we are capable of ensuring the application of privacy-enhancing mechanisms based on object types or object identities, such as hiding authenticated staff in an airport environment.

The implementation of such a privacy-aware workflow with respective enforcement mechanisms requires assumptions about the architecture of the target system. Resting on the work of Hampapur et al. [3] and Monari et al. [6], we derive a generic architecture, which serves the purpose of discussing and illustrating how usage control enforcement mechanisms have to be integrated into such systems.

We explicitly do not consider anonymization techniques for video data or abstracted object information. Such approaches are orthogonal and complementary to our work, since our usage control infrastructure can conveniently be leveraged for enforcing the application of privacy-enhancing mechanisms when visualizing, storing, and accessing data. Moreover, the utilization of anonymization techniques can hardly be reflected in a generic workflow, as the achievable gain for privacy as well as the appropriate degree of anonymization strongly depends on the context and on the requirements of the surveillance purpose (cf. §7).

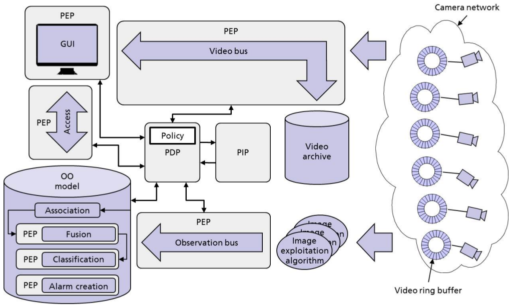

Figure 1. A generic architecture for modern video surveillance (darker components) augmented with usage control (lighter boxes)

We now outline a generic architecture for privacy-aware surveillance systems. Generalizing earlier work [3, 6], our abstraction subsumes many approaches to surveillance architectures in the literature.

To date, surveillance systems are neither technically capable to reliably work in a fully autonomous manner, nor do legal regulations of many legal spaces approve the deployment of systems taking automated decisions potentially producing legal effects, as for instance referred to in Article 15 of Directive 95/46/EC of the European Union. Thus modern surveillance systems support the operator through intelligent pre-processing of video streams, indicating noticeable events, and providing appropriate information and instruments for assessing and handling such incidents.

3.1. Cameras, Ring Buffers, and Video Archive

When pointing the operator to some incident requiring human assessment, the system needs to transmit video data. Cameras are often equipped with ring buffers (cf. Figure 1), which can be accessed to view the preceding few minutes concerning an incident under investigation. Many video surveillance systems also incorporate a video archive, which in the first place is used for preservation of evidence with regard to criminal prosecution.

3.2. Classes of Image Exploitation Algorithms

As illustrated in Figure 1, we assume that image exploitation algorithms are not yet fully integrated into video surveillance camera hardware. This is justified as algorithms like automated object tracking necessarily have to work on video streams of multiple cameras.

A surveillance system's image exploitation requirements strongly depend on the specific purpose of the surveillance measure. We argue that we can distinguish between two classes of image exploitation algorithms. The first class aims at detecting critical incidents and runs continuously, *e.g.*, to detect people falling down or to recognize left behind luggage in an airport. We refer to this class as *detective algorithms*. The second class of algorithms comprises more privacy-invasive surveillance operations, e.g. extracting biometric features, yet is only required on-demand, i.e. on behalf of a human operator while handling an incident. We call these algorithms *reactive algorithms*. We consider object tracking, possibly across cameras, to belong to the detective algorithms: Even though tracking in itself may be considered to be privacy-intrusive, object recognition and tracking is necessary if specific objects should be hidden from the operator's screen or excluded from further analysis. We thus have the seemingly paradoxical situation that tracking is *necessary* for protecting privacy. Note that the persistent storage of tracking data for further analysis, however, is considered to be reactive.

3.3. Object-Oriented Model of the Monitored Area

Complex tasks, such as recognizing left behind luggage, require an abstract representation of people, objects, and mutual relations in the monitored area. For this, output data of detective algorithms needs to be aggregated and consolidated. These algorithms send their detections to an observation bus, from where they are fed into a data processing chain, which establishes and maintains an object-oriented model of the observed area. Notable approaches have been proposed by Hampapur et al. [3] and Bauer et al. [1]. Data processing typically starts with an *association* step, in which the system determines whether an incoming observation refers to a known object (cf. Figure 1). The *fusion* step either aggregates the new information with an existing object or creates a new one. The *classification* step determines the type of the object given the new state of information. Depending on updated information, a previously unknown person is reclassified, *e.g.*, as a police officer.

3.4. Graphical User Interface

The distinction of video data and abstracted data is also reflected in the design of GUIs. Large video walls are eliminated by site map views of the monitored area using stylized renderings of persons or objects (*e.g.* pictographs), while only selectively visualizing video for situation assessment.

4. A Privacy-Aware Surveillance Workflow

Based on the generic architecture of §3, we introduce a workflow template that distinguishes two *operational modes*, reflecting the classes of image exploitation algorithms of §3.2: The *default mode* encapsulates *detective* algorithms, the *alarm mode* encapsulates *reactive* algorithms. Policies for both modes configure the policy decision point (PDP) in Figure 1.

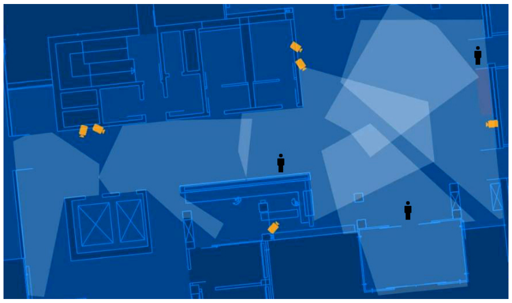

Figure 2. Site map view of the prototype system in *default mode*: people are visualized as pictographs, no video data is exposed

4.1. The Default Mode

In the *default mode*, the system executes *detective* image exploitation algorithms. As introduced in §3.2, such algorithms perform continuous tasks inherent to the purpose of the concrete surveillance measure in order to point the operator to noticeable incidents. While no incidents are detected, the only type of data aggregated in the object-oriented model is the current positions and types of objects. Thus the operator's GUI needs to merely show a site map view (cf. §3.4). In addition to stylized visualizations of persons and movable objects, this view includes pictographs of cameras and their fields of vision. Figure 2 shows our prototype system working in default mode. This mode is privacy-aware as it does neither grant access to video data, nor does it store any data persistently. However, there are limitations concerning achievable privacy (cf. §7).

Upon detecting an incident the workflow requests the operator's assessment of the situation. Depending on the operator's feedback, the workflow switches to the *alarm mode*. In order to enable investigations, an alarm needs to be associated with the person who set it off and the camera that delivered its trigger event. Establishing such associations may require interaction with the operator. Assume that a suspicious piece of luggage is detected. At detection time, the person who dropped the luggage object may already be gone. According to the latency of the detection of dropped objects, a certain time-frame is retrieved from the respective camera's ring buffer and searched for the object's owner. As the search result may be ambiguous, the operator needs to select a particular person from preview pictures.

4.2. The Alarm Mode

The *alarm mode* displays the video stream of the camera that triggered the alarm (Fig. 3). The person who provoked the alarm is highlighted in the video stream as well as on the site map. The operator is allowed to access arbitrary camera live streams and ring buffers (*e.g.*, by clicking on the camera pictographs). The operator may also make the system execute more privacy-invasive operations, *e.g.*, an-

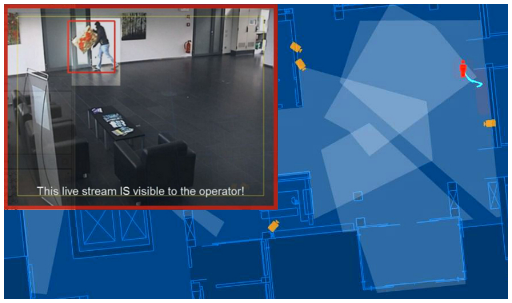

Figure 3. Site map view and live video stream in *alarm mode*: The surveillance system detects a person stealing a painting from an art exhibition

alyzing tracking data of a person, extracting biometric features, or recording video data for preservation of evidence. In return, the operator's interactions with the surveillance system are logged in detail. The alarm mode is either left by closing the incident as *resolved* or as a *false positive*. Data of resolved alarms including interaction logs is moved to a cryptographically secured storage. Given a false alarm, any related data is deleted besides the logs.

5. Usage Control Requirements in Modern Video Surveillance Systems

To ensure that the system is operated in compliance with the privacy-aware video surveillance workflow, we propose to deploy usage control enforcement. We derive UC requirements and point out where UC mechanisms are usefully integrated into our generic architecture of §3.

5.1. Usage control requirements of the default mode

We specify mode-based usage control policies that prohibit specific system operations in the default mode, and unlock them in the alarm mode. Such intrusive operations include access to cameras, recording video, analyzing tracking data, or extracting biometric features. The enforcement of such policies requires a first policy enforcement point (PEP) for the graphical user interface, intercepting the actions, which enable such system operations (cf. Figure 1).

Assuming that intelligent surveillance systems can classify objects, *e.g.*, airport staff and air passengers, the default mode also implements privacy-enhancing features based on object types (cf. §3.3). In type-based UC policies, we can for instance specify that airport staff must not be visualized within the site map view. This requires PEPs governing access to the object-oriented model and intercepting (re-) classification of objects, since a new object type may match a further applicable policy. Both cases, UC on individual data items as well as classes of items, can be catered to by the event-based interpretation of usage control (cf. §2).

5.2. Usage control requirements of the alarm mode

Since in alarm mode usually almost all functionality of the system is going to be operational (cf. §4.2), logging, encrypting and storing data is the predominant usage control requirement ("guarding the guards"). The enforcement can be performed by a specific PEP for the GUI. Note that logging user interactions does not prevent the operator from abusing the system; this abuse, however, can be detected.

In order to ensure selectivity while unlocking privacy-invasive operations in case of an alarm, such data, *e.g.*, biometric features, must only be collected from the particular person associated with the given alarm (cf. §4.2). Since image exploitation algorithms do not know about objects and alarms, two complementary PEPs are required for enforcing alarm-based usage control policies. A PEP on the level of the observation bus (cf. §3.3) tags detections with an ID of the actual alarm scope. The second PEP intercepts the fusion of data into objects. Combining both PEPs, we can ensure information flow conditions such as "biometric face templates must only be fused into objects that are associated to the same alarm scope in which their collection has been triggered." To do so, we combine usage control enforcement with information flow tracking presented in §2, which, generally speaking, makes it possible to disallow arbitrary aggregations. We instrument a policy information point (PIP, cf. Figure 1) for keeping track of the binding between sensitive data and the corresponding alarm scope.

Video data is handled analogously. Assume that the operator enables recording of video data for perpetuating evidence concerning a given alarm. The PIP again keeps track of the binding between chunks of video data and the alarm scope in which their recording has been initiated.

Encapsulating privacy-sensitive data into the scopes of alarms becomes relevant when alarm handling is finished. Then, as explained in §4.2, additional data collected in the scope of the alarm must either be encrypted for preservation of evidence, or deleted in case of a false alarm (except for logs for which archiving is mandatory). Either way, the PIP provides the knowledge about the particular pieces of data belonging to the scope of a given alarm. PEPs for the video archive and for the object-oriented model provide mechanisms for deleting data or ensuring encrypted archiving.

6. Implementation

We chose two scenarios for demonstrating usage control enforcement in the video surveillance testbed NEST [6].

The first scenario is about detecting theft of paintings in an art exhibition. We only provide a site map view in default mode. If a painting is moved, the alarm mode is entered and the live stream of the associated camera is visualized. Access to other cameras as well as locating and displaying the potential thief on the overview map is allowed. Usage control on the level of the GUI governs interactions with the operator and visualization functions. Usage control on the object-oriented model allows to restrict the analysis of tracking data to the person associated with the given alarm.

In our second scenario, we enforce type-based policies, *e.g.*, we hide persons authenticated as employees in the default mode. This requires a two-stage authentication with the surveillance system using a mobile communication device, *e.g.*, a smart phone or tablet [13]. First, a cryptographic authentication is performed over a wireless network, authenticating the mobile device as belonging to an employee. In the second step the system replies with a short-lived graphical code, which is easy to recognize for surveillance cameras. When the code is presented to a camera, the authentication identity as an employee is fused into the associated *unknown person* object. The object's type is hence reclassified to *employee*. The association of an object and its (group) identity is maintained by employing the system's tracking capabilities. We implement usage control at the level of classification by intercepting object classification events. These events trigger mechanisms at the level of the GUI, which prohibit object visualization.

7. Limitations and Conclusions

Limitations. Despite not disclosing video data, the site map view (cf. §3.4, §4.1) may not be sufficient to protect the privacy of people in the monitored area. This is because operators may have additional context knowledge. Assume a video surveillance measure in a hospital. The purpose is to support night nurses by detecting patients falling down or wandering about in corridors. Knowing the location of the nurses' room combined with the hospital's duty roster allows for creating movement profiles of individual nurses, which can be abused for performance monitoring, *e.g.*, assessing a nurse's reaction time in case of paging patients. Similar threats exist for image anonymization techniques: Blurring out faces fails to protect a person that can be identified due to distinctive clothing, *e.g.*, a uniform. A silhouette view does not preserve the privacy of a person walking on crutches, or performing some specific task. [10] constitutes a first step towards modeling such context knowledge.

Media breaks are beyond the scope of usage control mechanisms. If a malicious user films screens in alarm mode, the respective movie is not protected.

Conclusions. To increase privacy in the face of expanding video surveillance, we have presented a perspective towards deploying usage control in modern surveillance systems. We derived key usage control requirements of privacy-aware workflows and demonstrated how according usage control mechanisms can be implemented within a generic video surveillance architecture.

By showing how to implement ideas such as restricting the utilization of privacy-invasive operations to a specific operational mode as well as encapsulating collected data into scopes of alarms, we have shed light on the potential of transferring the conceptual framework of usage control to the domain of privacy-respecting video surveillance.

References

[1] A. Bauer, T. Emter, H. Vagts, and J. Beyerer. Object oriented world model for surveillance systems. In *Future security: 4th Security Research Conference*. Fraunhofer Verlag, 2009.

[2] D. A. Fidaleo, H.-A. Nguyen, and M. Trivedi. The networked sensor tapestry (NeST): a privacy enhanced software architecture for interactive analysis of data in video-sensor networks. In *Proc. 2nd ACM intl. workshop on Video surveillance & sensor networks*, pages 46–53, 2004.

[3] A. Hampapur, L. Brown, J. Connell, A. Ekin, N. Haas, M. Lu, H. Merkl, and S. Pankanti. Smart video surveillance: exploring the concept of multiscale spatiotemporal tracking. *IEEE Signal Proc. Mag.*, 22(2):38–51, 2005.

[4] M. Harvan and A. Pretschner. State-based usage control enforcement with data flow tracking using system call interposition. In *3rd Intl. Conf. on Network and System Security*, pages 373–380, 2009.

[5] F. Kelbert and A. Pretschner. Data usage control enforcement in distributed systems. In *CODASPY*, pages 71–82, 2013.

[6] E. Monari, S. Voth, and K. Kroschel. An object- and task-oriented architecture for automated video surveillance in distributed sensor networks. In *Advanced Video and Signal Based Surveillance, 2008. AVSS '08. IEEE Fifth International Conference on*, pages 339–346, Sept. 2008.

[7] J. Mossgraber, F. Reinert, and H. Vagts. An architecture for a task-oriented surveillance system: A service- and event-based approach. In *5th Intl. Conf. on Systems*, pages 146–151, Apr. 2010.

[8] A. Pretschner, M. Hilty, and D. Basin. Distributed usage control. *Commun. ACM*, 49(9):39–44, Sept. 2006.

[9] A. Pretschner, E. Lovat, and M. Büchler. Representation-independent data usage control. In *Proc. DPM/SETOP*, pages 122–140, 2011.

[10] M. Saini, P. K. Altrey, S. Mehrotra, and M. Kankanhalli. W^3-privacy: understanding what, when, and where inference channels in multi-camera surveillance video. *Multimedia Tools and Applications*, Aug. 2012.

[11] A. Senior, S. Pankanti, A. Hampapur, L. Brown, Y.-L. Tian, A. Ekin, J. Connell, C. F. Shu, and M. Lu. Enabling video privacy through computer vision. *Security & Privacy, IEEE*, 3(3):50–57, May-June 2005.

[12] H. Vagts and A. Bauer. Privacy-aware object representation for surveillance systems. In *7th IEEE Intl. Conf. on Advanced Video and Signal Based Surv.*, pages 601–608, 2010.

[13] H. Vagts and J. Beyerer. Enhancing the acceptance of technology for civil security and surveillance by using privacy enhancing technologies. In *Future Security*, pages 372–379. Fraunhofer, 2011.

[14] J. Wickramasuriya, M. Datt, S. Mehrotra, and N. Venkatasubramanian. Privacy protecting data collection in media spaces. In *Proc. 12th annual ACM intl. conf. on Multimedia*, pages 48–55, 2004.