Identifying and Addressing Challenges for Search and Analysis of Disparate Surveillance Video Archives

Suzanne Little*, Kathy Clawson†, Anna Mereu‡, Aitor Rodriguez§

*Dublin City University, Ireland, suzanne.little@dcu.ie
†University of Ulster, United Kingdom, k.clawson@ulster.ac.uk
‡HI Iberia, Spain, amereu@hi-iberia.es
§IKUSI, Spain, aitor.rodriguez@ikusi.com

Keywords: surveillance video, search, analysis

Abstract

This paper discusses the challenges faced when bringing together multiple disparate surveillance video archives to support semantic analysis and search and describes the SAVASA framework for enabling better integration of CCTV archives. The proliferation of CCTV cameras managed by public institutions and private enterprise raises a number of issues relating to data security, privacy, ethics and technological difficulties in unifying the variety of data and formats. These are often the result of misunderstandings between the involved parties and include concerns such as reliability and security of cloud technology, accuracy and privacy in automatic semantic annotation of video and ultimate responsibility over content in the digital age. In this paper we present outcomes from the SAVASA project that aims to develop a standards-based video archive search platform allowing authorised users to query over remote and non-interoperable video archives of CCTV footage from geographically diverse locations.

1 Accessing disparate surveillance video archives

The increasing ubiquity of CCTV and surveillance video systems results in very large archives of footage captured and recorded in remote locations, at different levels of coverage and with different formats, available metadata or searchable indices. Authorised users face many challenges accessing specific footage or finding relevant segments based on semantic descriptions such as "white car", "person running". Both research and commercial efforts have sought to improve automatic analysis, management and information extraction from digital video footage [10, 8].

The SAVASA[1] project – Standards-based Approach to Video Archive Search and Analysis – aims to develop a standards-based video archive search platform that allows authorised users to query over various remote and non-interoperable video archives of CCTV footage from geographically diverse locations. At the core of the search interface is the application of algorithms for person/object detection and tracking, activity detection and scenario recognition. The project also includes research into interoperable standards for surveillance video, discussion of the legal, ethical and privacy issues and how to effectively leverage cloud computing infrastructures in these applications. Project partners come from a number of different European countries and include technical and research institutions as well as end user, security and legal partners.

Recent developments in CCTV systems have resulted in higher resolutions and new technologies to can take advantage of analytical algorithms that monitor footage and provide alarms or automated actions through previously setup triggers. Furthermore, CCTV devices are becoming Edge devices (entry points), storage systems and/or off-the-shelf servers instead of proprietary DVR (as in the past). This means video processing can be fully integrated and supervised by software systems.

This paper looks at the challenges of providing integrated search of surveillance video archives from different locations focussing on how we addressed three particular issues: security and data control; privacy and data protection and support for rich semantics. A sample of the current video retrieval procedure is described and two use cases are presented to outline the type of applications envisaged. Several related projects and commercial offerings are identified. The SAVASA framework is described and the main developments designed to address concerns raised by end users are discussed. Finally the remaining challenges and future direction of the project are presented. The contribution of this paper is in identifying some of the overlooked concerns and practical implications of applying cutting edge technology for CCTV analysis and it is hoped that this can help bridge the gap between the motivations and expectations of both researchers and end users.

1.1 Use cases

This section describes an example of how video is currently retrieved by law enforcement agencies (LEA) and provides two use cases targeted by the SAVASA project to illustrate the application of the video archive integration and search functionality.

Most of the camera data is locally recorded and stored on

[1] `http://savasa.eu`

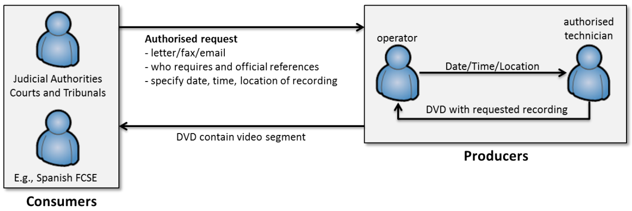

Figure 1. Example of current process for retrieving CCTV

hard disks. The video archives must be deleted on or before a defined storage duration, e.g., 30 days according to the Spanish LOPD (Data Protection Law). Currently, this regulation varies between European countries. Old recordings are replaced by new ones according to available storage and existing regulations regarding data retention periods.

Figure 1 illustrates an example process for retrieving CCTV footage. An official authorised request is needed in order to allow the extraction of video images. This extraction is done by an authorised technician who goes physically to the video recorder where the images are stored and copies them to a DVD. The software for playing the video is automatically saved with the video archive because the resulting file can be in a unique or proprietary format. These formats or extensions depend on the camera/recorder model.

Watermarking is applied on all the recordings just after the video signal is captured. The watermarking algorithm is usually different for each recorder model or CCTV solution provider. The resulting recordings of CCTV systems are declared suitable for security purposes, so that they can only be displayed if they are officially required by the courts, local or national security forces. In these cases, videos constitute a basis for investigations of events such as theft, accident investigation, checking alibis, organized crime and to investigate possible terrorist attacks, among others.

The main benefit offered by CCTV systems to enterprises is related to the monitoring in real time of the events, often reported by private security facilities, in order to remotely manage the involvement of more staff, enforcement personnel or health services. Two current use cases under investigation within SAVASA are described here:

Person falling A person presents a claim to have fallen at the entrance to a train station. Following the incident and with the purpose of investigating the location of the fall for insurance purposes, a judge issues an order to obtain a copy of the images from all cameras at the station entrance. The Civil Protection and Safety Direction, receives the order, signed and stamped. Once the extraction is approved by the data controller (operator), a work order is issued and sent to the technical staff. The staff goes to the station and removes the video (taking into account the cameras, dates, and requested slots of time). Images are stored on a DVD, and a duly identified and authorized person (usually a police officer or civil guard) picks up the images, signing a document as proof of the delivery.

Crowd fighting After a football match, at the train station there is a violent confrontation between the fans of both teams. The altercation ended with the arrest of some of those involved. In order to clarify the causes, and to facilitate investigations of the participation of detainees, a court order is received and verified by the operator. Once the extraction is authorized, a work order is sent to the technical team who makes a DVD copy of the requested images, which will be hand-delivered to maintain the appropriate chain of custody.

1.2 Challenges

There are many ongoing challenges in automated and integrated surveillance systems [4]. The key barriers that must be overcome to ensure timely, accurate and legal access to surveillance footage are:

- harmonisation of metadata standards and semantic terms to enable search over multiple archives;
- indexing of video at sufficient granularity (e.g., person tracking, object detection, time stamping etc.);
- support for integrated and remote access to databases;
- foundational integration of legal, ethical and privacy concepts (‘privacy-by-design’ [6]) and
- strong security support.

These issues are well known and section 2 outlines a number of research projects and commercial offerings that seek to provide solutions. However, in discussions with end user partners, SAVASA researchers discovered new insights into the level of importance placed on some concerns and other challenges not generally considered.

When first presented with the concept of an integrated, cloud-based search solution for geographically distributed CCTV archives, end users were particularly uncertain about the transfer of video data through networks outside of their control, especially where transport was labelled as via ‘The Internet’. There were very strong reservations about the security of the cloud, who would maintain legal ownership and responsibility of the data and ultimately who would have both public and legal responsibility for any breaches in security or data leakage. Existing corporate regulations forbidding the transfer or long-term storage of video would also prevent the type of sophisticated, remote and semantically-rich video search that had been envisioned. Section 4.1 discusses how the framework was adapted to mitigate some of these risks and to ameliorate the perception of the cloud infrastructure.

An unexpected difficulty also arose due to the reluctance or inability of video producers to make their images available outside of their networks. Exploratory or faceted search methods are considered useful interaction models for surveillance systems – two recent examples can be found in [2, 5]. The basis of exploratory, iterative or browsing-based search models requires feedback from the user who considers the provided search results and adjusts their query accordingly [9]. For example, a user might first narrow down a results set by date, then by number of people in the scene and then filter by attributes such as colour, motion, position. Without viewing the resulting videos it is impossible for a user to determine if the specific occurrence they are looking for has been found. If thorough, complete and very accurate semantic annotations are available and the user is confident in the accuracy of the metadata then search for a specific event by text alone may be feasible.

Not only is metadata accuracy required for successful search but the perception of automatic semantic annotation capability and search functionality has a strong impact on the users' overall impressions of the project as we found through participating in the TRECVid Interactive Surveillance Event Detection task [7]. Users' expectations were both realistic and somewhat flawed. Successful commercial systems, highly trained and deployed in vertical applications, could lead to inflated expectations of what is generally achievable by computer systems. However, users were also somewhat jaded by previous experience in seeing systems fail when applied to their specific requirements. In addition, research projects in the surveillance domain, face considerable difficulty in collecting sufficient annotated example video footage to conduct large-scale development and evaluations. This is due, in large part, to privacy and data protection regulations and the difficulty in gaining permission to gather data from real-world systems but also to the general problem of building training data in computer vision research. Tackling this challenge requires ongoing communication, understanding and management of expectations. Section 4.2 outlines some of the methods used to improve the adaptability of the semantic framework.

The users' concerns about data control and usable semantic search led to adjustments in the design of framework as part of the project's iterative development process. Section 3 describes the current version of the framework, however there are many relevant projects that tackle different parts of these challenges and they are described in the next section.

3DTVS (2011–2013) 3DTV AV content description, indexing, search and browsing across open platforms. `http://www.3dtvs-project.eu`
ADDPRIV (2011–2014) Platform for automatic semantic content discrimination within video surveillance data. `http://www.addpriv.eu`
ADVISE (2012–2015) Focus on design and development of a unification framework for surveillance-footage archive systems. `http://www.advise-project.eu`
IMPART (2012–2015) Intelligent Management Platform for Advanced Real-Time media processes. `http://impart.upf.edu`
MOSAIC (2011–2014) Multi-modal data intelligence capture and analytics including video and text collaterals. `http://www.mosaic-fp7.eu`
MUSCLE (2005–2008) Research project exploring the full potential of statistical learning and cross-modal interaction for the (semi-) automatic generation of semantic meta-data for multimedia.
SAMURAI (2008–2011) Suspicious and abnormal behaviour monitoring using a network of cameras & sensors for situation awareness enhancement
SUBITO (2009–2011) Surveillance of Unattended Baggage and the Identification and Tracking of the Owner
SV3D (2011–2013) Surveillance platform based on multi-source video analytics, localized data and cognitive interfaces. `http://fp7-sv3d.com`
VANAHEIM (2010–2013) Integration of innovative audio/video analysis tools within a CCTV surveillance platforms typically in use in urban transport environments (metro and railway stations). `http://www.vanaheim-project.eu`
VideoSense (2011–2015) Virtual Centre of Excellence for Ethically-guided and Privacy-respecting Video Analytics `http://www.videosense.eu`

Table 1. Surveillance and multimedia projects

2 Related Work

A number of related projects and commercial offerings focus on research and application areas related to surveillance video integration, analysis or search. Table 1 lists a number of the most relevant recent research projects.

In addition to innovations generated through research projects, commercial options are available for specific applications relating to surveillance video analysis. A number of companies offer video recording, analytics and management services including Geutebruck[2], Bosch CCTV[3], IProNet[4], Scati Video Management Systems[5], Wavestore[6], and Intellio[7]. Semantic tools include face recognition, person detection and counting, object detection, people tracking, license plate recognition, and intruder detection. Content management and administrative functionalities include security information management, remote setup, user management, and hardware health monitoring and configuration. Despite the large range of commercial analytics and management services available, such offerings are generally restricted to a particular installation and commercial hardware. Furthermore, the environmental constraints under which specific methodologies are effective are not always clearly cited and thus competing methods can be difficult to compare in terms of performance.

Fusion of inputs from multimodal data is a technology area in which major advances are expected. The idea is to integrate the information captured by the cameras with data offered by other type of sensors such as combining audio and video [3]. Thus, the information can be fused to create a visualization of a more complete "big picture" of the area being surveilled. For example, in the SAMURAI European project based on the information provided by a heterogeneous set of sensors, an innovative intelligent surveillance system for robust monitoring of both inside and surrounding areas of a critical public infrastructure site has been developed.

The automatic detection of abandoned equipment is an example of another specific area of active investigation. The SUBITO project focuses on the automated real time detection of abandoned luggage or goods and the fast identification of the

[2] `http://www.geutebruck.com/index.php?Culture=en_EN&PageId=454/`

[3] `http://www.boschsecurity.com`

[4] `http://ipronet.es/?language=english`

[5] `http://www.scati.com`

[6] `http://www.wavestore.com/`

[7] `http://intellio.eu/en/`

individual who left them and their subsequent path. The goal is that video surveillance systems deployed in places such as airports and train stations become capable of identifying baggage that has been left unattended, and that could therefore possibly contain an explosive device. Then it will be searched the identity of the person who has deposited the baggage, to follow it through various cameras and establish its present location. Finally, new trends in CCTV technologies are integrating the analytical process within the camera. In the near future it is hoped to transmit only alarms, with playback or live video as needed rather than a complete stream.

The projects and commercial products listed above, provide some excellent solutions and exciting future possibilities to improve capture, analysis and access to surveillance video networks. However, as noted by research firm IMS/IHS in successive reports on the state of the security industry[8], there remains resistance to integrated solutions and the use of technologies such as cloud infrastructure, Video Surveillance as a Service (VSaaS) and lack of confidence in visual semantic annotation services and the ability of the operator to maintain adherence with ever shifting data protection requirements. The next section presents the SAVASA framework and discusses the evolutions that have taken place in response to user feedback.

3 The SAVASA Framework

The SAVASA platform will allow authorized users to perform remote semantic queries on different interoperable video files in an effective way. To achieve this goal it has been necessary to deepen the concept of Video Surveillance as a Service (VSaaS) and propose an architecture that enables secure remote access to the semantic search functionalities provided by the platform. As shown in Figure 2, the core of the architecture is the cloud infrastructure (hosting different modules) that stores and manages all the data associated with the videos and efficiently delivers it on-demand to the users.

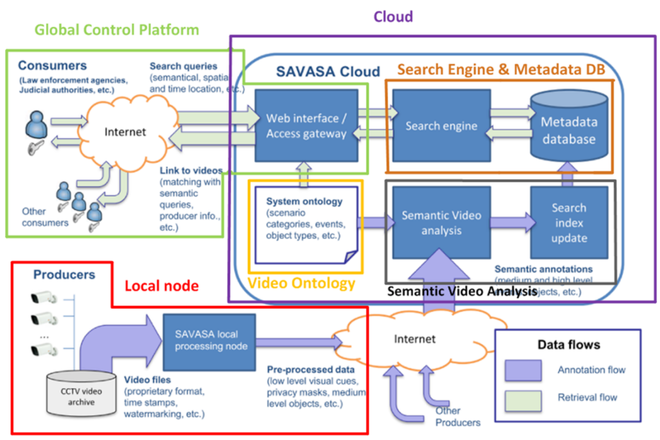

Figure 2. Modules of the SAVASA framework

In this figure the two types of users who can access to SAVASA are clearly defined. On one side are the *Producers*, which are those transport entities that generate and manage the video files. These users "feed" the platform database with the information provided by their video surveillance systems. RENFE, the main Spanish railway operator, along with other European transport entities has fulfilled this role in the project. On the other side are the *Consumers* or Law Enforcement Agencies (LEAs), who realize searches for different events, for example a fight at a train station, to get the video recordings that help in their police investigations.

SAVASA not only offers a search and analysis tool for the CCTV files but also ensures that access to them is made according to legal and ethical guidelines established by the European Union for the protection of personal data. With aim that videos never leave the premises of the Producers, the local node has been separated from the other Cloud modules. This node is installed in the facilities of the Producer and accesses the videos to perform image processing, extract low-level visual features, as well as perform the detection and tracking of objects of interest like people or vehicles. The pre-processed data are uploaded to the Cloud SAVASA through secure VPN connections. The modules that are part of the platform are the following:

Local processing node Local nodes will be deployed in the Consumer facilities. These nodes will convert CCTV videos to a common non-proprietary format and also will pre-process data before transferring information to Cloud-based services.

Global control platform Global Control Platform provides the SAVASAs user interface for consumers and producers. This platform will also store the web services for access and outputs to SAVASA as well as for the interoperability with other modules.

Search Engine and Metadata Database Search Engine will find in the Metadata Database the annotations that best match to the user queries. The results list contains sufficient information to identify the video and the segment or region that has been labelled with the required object or event.

Video Ontology Video ontology will be used to expand or refine queries defined in the user application and performed by the Search Engine. It is an ontology which describes concepts related to video content which are relevant for the SAVASA surveillance domain.

Semantic Video Analysis Semantic video annotation will take non-identifying low-level feature descriptors extracted from video and generate annotations describing the objects and events defined by the ontology. This may include people, vehicles, general events (e.g., person runs, car accident). The goal is to create a search index based on the annotations provided by this module where possible without requiring the original source video to be present.

[8] http://www.securitymarketintelligence.com/press_releases/Video_Surveillance_Trends_for_2013, Accessed 2013-09-12

Cloud The set of modules that provides the main functionalities of the SAVASA will be deployed in a Cloud infrastructure. The cloud environment of SAVASA introduces a novel cloud service; namely Video Surveillance as a Service (VSaaS), which is specifically tailored to Video Surveillance requirements providing interoperability, centrally controlled secure remote access, unified indexing system and data storage.

4 Addressing the Challenges

4.1 Security and Data Control

As discussed in section 1.2 there was significant resistance to the initial concept of transferring video through the Internet to cloud-based systems and out of the control of the video producers. To address these concerns, two main changes were made to the framework involving the addition of the local node (seen in Figure 2). The local node enables initial processing, transcoding, privacy masking and watermarking of the video to occur prior to any options for indexing, video access or transfer. In scenarios where the video data is not permitted to be transferred, semantic annotation and indexing is performed using low-level visual feature descriptors that can be exploited by trained classifiers but do not provide any possibility of reconstructing the video file.

Concerning remote access from Consumers and Local Nodes to the "core" of the cloud, SAVASA offers secure connections through Virtual Private Network (VPN), using two alternative access technologies. The connection of Local Nodes is established through gateway-based connections, while the Consumers connect to the platform through a VPN client software. Figure 3 illustrates the gateway connection used to integrate the Local Node.

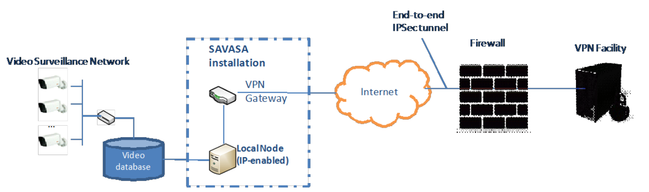

Figure 3. Gateway connection for local node within cloud

SAVASA's cloud platform adopts the open source OpenStack operating system [1] that offers scalable virtual networks for interconnecting Virtual Machines (VMs). The SAVASA cloud is configured as a single cloud domain, realizing a Cloud Controller, several Computer Nodes for Virtual Machines (VMs) deployment and a Storage Node Controller for storage support. The extension of this platform is easily achievable in order to support various loads and applications.

Privacy-by-design [6] is an integral part of all security-related research projects. Therefore a key consideration in discussions about the SAVASA project was ensuring that existing processes for managing access to surveillance video data were respected. The addition of the local node helps to support data ownership and protection but provides the structure to support better integration of data archives where the necessary legal authorisation has been received.

4.2 Integration and Semantic Support Framework

Search over massive amounts of videos in the Video Surveillance domain presents the challenge of dealing with different video formats as well as different collections of metadata associated to it which then makes it difficult to develop interoperable search solutions. In situations where heterogeneous video archives sources have to be integrated and searched through in a seamless way, this process might very complex and often time consuming if the multiple archive system does not support interoperability.

Semantic Web Technologies allow video data representation and description through Ontologies which constitute a formal explicit description of concepts, or classes in a specific domain, in this case Video Surveillance, and the relationships that relate them among each other. The Ontologies allow describing concepts and relationships by enabling overcoming the heterogeneity of video resources and video archive search and exploration using standard base knowledge interchange technologies.

The effort of describing the SAVASA Video resources content by means of concept hierarchies and relationships enables bridging the gap among variety of video formats and resources and the heterogeneity of search methodologies and paradigms that could be deployed in the SAVASA Platform. The Video Ontology is a data description format that enables describing data highlighting concepts to which the specific piece of data can be related to and the relationships it has with others data that may be described according to the ontology structure. The SAVASA Video Ontology is a component that is important within the Semantic Video Analysis and Annotation tools as it provides a coherent description of the Video Surveillance concepts and relationship that are detected in the Videos and provide input for Videos labelling.

Figure 4 depicts the interaction of the Semantic Support Framework with the other modules of the SAVASA framework.

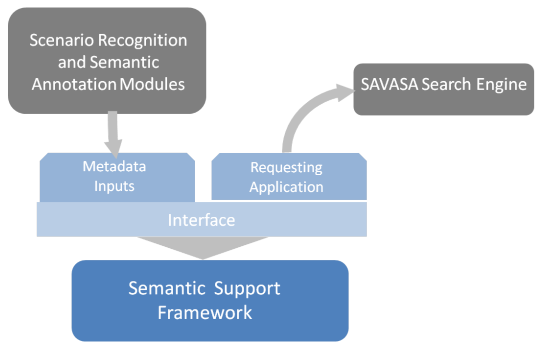

Figure 4. Semantic Support and Integration Framework

The Scenario Recognition and Semantic Annotation Modules are able to produce metadata for the processed videos which describe the set of events that have been detected and categorized according to the concepts defined in the Video Ontology. These metadata have to be processed to be compliant to the metadata defined in the Ontology and inserted in the repository. From the other hand, the SAVASA Search Engine will perform queries using the Ontology in order to gather specific events of interested from the stored events set. In order to do so, the Semantic Support Framework exposes a set of REST Interfaces to allow the integration with such modules.

This methodology is provided in the SAVASA Framework in the following way: the Semantic Video Analysis and Search Index update provides the video annotations according to a common understanding based on the developed Ontology. At search stage, search concepts and relationships belonging to the Ontology are retrieved by means of semantic queries that are processed over the Semantic Annotation repository (the metadata database) performing the information retrieval flow. For this purpose high level APIs are being developed to mediate between the low level video analysis stage and the high level of video search.

5 Conclusions and Future Challenges

The SAVASA framework is designed to support better integration of disparate video archives to improve the efficiency of CCTV operators in responding to official data requests or enable expanded search for LEAs when permitted. The architecture provides strong security, data protection and integrated semantics support. The iterative improvements to the framework via the introduction of the local processing node have begun to address some of the concerns of key stakeholders, namely the control of access to their video data and removing the need to transfer video through unsecured networks.

Challenges still remain. In particular difficulty gaining access to real video data due to privacy concerns when recording in public places and how to evaluate "real" performance of the system in small scale installations. Specifically how to identify unknown and unexpected events. The framework is currently being formed into an integrated prototype, combining research from different partners, to conduct further testing. A key aim is to continue educating, communicating with and learning from end user partners, ethical and legal experts and other interested stakeholders.

This paper has identified some often overlooked concerns and practical implications of applying cutting edge technology for CCTV analysis and discussed how the SAVASA framework is seeking to address and evaluate technologies to allow users to query over disparate video archives. The understanding gained from this work will have broader implications for improving communication and understanding the motivations of both researchers and end users.

Acknowledgements

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement number 285621, project titled SAVASA.

Part of the content of this paper is from SAVASA project documents contributed to by the following persons: Francesca Gaudino (Baker & McKenzie), Roberto Giménez (Hi Iberia), Emmanouil Kafetzakis (NCSRD "Demokritos"), Tasos Kourtis (NCSRD "Demokritos"), Jun Liu (University of Ulster), Ana Martínez Llorens (RENFE), Giorgio Montefiore (SINTEL Italia), Marcos Nieto (Vicomtech), Noel E. O'Connor (Dublin City University), Karina Villarroel Peniza (RENFE), Celso Prados (INECO), Pedro Sanchez (IKUSI), Alan F. Smeaton (Dublin City University), Raúl Santos de la Cámara (Hi Iberia), Bryan Scotney (University of Ulster), Hui Wang (University of Ulster)

The SAVASA project gratefully acknowledges the valuable contribution of our end user partners.

References

[1] Openstack – open source software for building private and public clouds. `http://www.openstack.org/`, accessed 2013-09-12.

[2] G. D. Castanon, A. L. Caron, V. Saligrama, and P. Jodoin. Exploratory search of long surveillance videos. In *Proceedings of the 20th ACM international conference on Multimedia*, pages 309–318, 2012.

[3] M. Cristani, M. Bicego, and V. Murino. Audio-visual event recognition in surveillance video sequences. *Multimedia, IEEE Transactions on*, 9(2):257–267, 2007.

[4] A. R. Dick and M. J. Brooks. Issues in automated visual surveillance. In *International conference on digital image computing: techniques and applications*, 2003.

[5] M. Hoferlin, B. Hoferlin, G. Heidemann, and D. Weiskopf. Interactive schematic summaries for faceted exploration of surveillance video. 2013.

[6] M. Langheinrich. Privacy by design–principles of privacy-aware ubiquitous systems. In *Ubicomp 2001: Ubiquitous Computing*, pages 273–291, 2001.

[7] S. Little, I. Jargalsaikhan, K. Clawson, M. Nieto, H. Li, C. Direkoglu, N. E. O'Connor, A. F. Smeaton, B. Scotney, H. Wang, and J. Liu. An information retrieval approach to identifying infrequent events in surveillance video. In *Proceedings of the 3rd ACM International Conference on Multimedia Retrieval*, pages 223–230, 2013.

[8] Y. Liu, J. Fan, and D. Wang. Review of intelligent video surveillance with single camera. In *Fourth International Conference on Machine Vision (ICMV 11)*, pages 834923–834923, 2012.

[9] G. Marchionini. Exploratory search: from finding to understanding. *Communications of the ACM*, 49(4):41–46, 2006.

[10] X. Wang. Intelligent multi-camera video surveillance: A review. *Pattern recognition letters*, 2012.