Intelligent video surveillance

Rita Cucchiara[1], Andrea Prati[2] & Roberto Vezzani[1]
[1]*Department of Information Engineering, University of Modena and Reggio Emilia, Italy*
[2]*Department of Engineering Science and Methods, University of Modena and Reggio Emilia, Italy*

Abstract

Safety and security reasons are pushing the growth of surveillance systems, for both prevention and forensic tasks. Unfortunately, most of the installed systems have recording capability only, with quality so poor that makes them completely unhelpful. This chapter will introduce the concepts of modern systems for Intelligent Video Surveillance (IVS), with the claim of providing neither a complete treatment nor a technical description of this topic but of representing a simple and concise panorama of the motivations, components, and trends of these systems. Different from CCTV systems, IVS should be able, for instance, to monitor people in public areas and smart homes, to control urban traffic, and to identity assessment for security and safety of critical infrastructure.

Keywords: Video Surveillance, Object Detection and Tracking

1 Introduction

This chapter will introduce the concepts of modern systems for Intelligent Video Surveillance (IVS), with the claim of providing neither a complete treatment nor a technical description of this topic, but of representing a simple and concise panorama of the motivations, components and trends of these systems.

There exist several definitions of the term *surveillance*. Among them, one interesting definition is that *surveillance concerns models, techniques, and systems for acquiring information about the 3-D external world, detecting targets along the time and space, recognizing interesting or dangerous situations, generating real-time alarms recording meaningful data about the controlled scene* [1]. The explosion of requests of surveillance systems for *protecting critical infrastructures* is due to the need for security that has followed recent catastrophic events, such as September 11, 2001. These events have contributed to the dissemination of advanced technologies, including

WIT Transactions on State of the Art in Science and Engineering, Vol 54, © 2012 WIT Press
www.witpress.com, ISSN 1755-8336 (on-line)
doi:10.2495/978-1-84564-562-5/11

video surveillance, able, for instance, to monitor people in public areas and smart homes, to control urban traffic, and to assess people identity for security and safety [2]. Recently, the term *video analytics* has also been extensively used to indicate the capability of automatically analyzing videos to detect and determine temporal events.

Thanks to the nature and richness of the provided information, cameras are widespread as the most used devices for monitoring and protecting critical infrastructures. In recent years, there has been a profound change in the way cameras and video surveillance have been used, moving from passive instruments for monitoring and storing images to *active tools for the prevention and the prompt reaction* in the case of abnormal or interesting events, leading to the so-called IVS. Until a few years ago, in fact, the major control centers of the police worldwide used the multitude of cameras only to monitor the various sites of interest and to manually detect areas at risk or events of interest, thus assisting the action of colleagues on site. This enormous (and often unnecessary) amount of information was then stored and used for criminal investigations and forensic activity in support of judges, with serious problems, however, of information retrieval (the typical "needle in the haystack"). It has been estimated that the continued presence of a human observer in private U.S. agencies costs approximately $150,000 per year. Finally, several studies in the psychology of perception agree that human attention degrades below the limit of acceptability after only 22 minutes of active video monitoring [3].

The recent advanced systems for video surveillance are often referred as *multimedia surveillance systems* [4]. The adjective *multimedia* is normally referred to systems and services conceived for human end users for accessing and using multimedia data. Applied to video surveillance, this term does not refer only to a system with the output in a multimedia format, but this system is also capable to collect, process in real time, correlate, and handle multimedia data coming from different sources (image, video, sensors, audio, etc.).

The leaning toward active and intelligent tools has dramatically increased the research activity in video surveillance over the past 5 years. In fact, the development of hardware and software solutions in complex and real scenarios is far from being straightforward. In addition, the diffusion of cameras over the major cities of the world to protect critical infrastructures calls for the use of *active multicamera systems* [5], which exploit distributed multiple cameras to provide redundant information and different viewpoints, even if they pose architectural and development issues.

Thus, a real-world IVS application usually requires a good integration and replication of a large plethora of modules (which will be detailed in the next section): *integration* in order to face compound problems and *replication* to manage more than one video source at the same time. Research interests have thus migrated from simple integration calls for specific solutions for handling camera handoff (to pass tracked objects between cameras' fields of view), methods for determining the best view given the scene's context, and sensor-fusion algorithms to best exploit a given sensor or sensor modality's strengths.

Leading companies in surveillance have proposed their own integrated frameworks, specially focusing on reliability, extendibility, and scalability aspects. For example, IBM S3 [6] is an open and extensible framework for performing real-time event analysis. New detection capabilities can be included by adding plug-ins to the low-level system (SSE, Smart Surveillance Engine), while all the fusion tasks are provided at the high level (MILS, Middleware for Large-Scale Surveillance). Similarly, ObjectVideo proposed the VEW system [7], developed starting from the prototype made under the VSAM (Video Surveillance and Monitoring) program [8]. Sarnoff Corporation proposed a fully integrated system called *Sentient Environment*, which can detect and track multiple humans over a wide area using a network of stereo cameras.

The ideal IVS system should have the following requirements:

- *Security*: The system should be capable to work also in challenging environments (due, for instance, to weather conditions or heat). It should also provide trustful data, with a certain guarantee of authenticity and protection against tampering, video manipulation, and so on.
- *Privacy*: Accessibility to raw video feeds is the way of disseminating data from camera systems. In some cases, live videos are streamed directly on the web, posing serious problems regarding privacy. Even in the case of video data directly accessible from authorized operators, potential misuses of these data are possible, and more dedicated solutions to protect privacy should be adopted [9].
- *No overstraining data*: As above mentioned, operators are overcommitted to watch tens of cameras at one time and are easily distracted by this huge amount of information (not to mention the distraction given by PTZ—Pan Tilt Zoom—cameras). Keeping a single person tracked in a single camera or even when moving between cameras can be a very challenging operation.
- *Flexibility*: The system must be designed to be easy to extend with new cameras and sensors, and with new functionalities. It should also be pluggable onto existing infrastructures, both in terms of cameras, data communication, and processing power.

The following section of this chapter will briefly introduce the main components of the architecture of an IVS system and report some examples of applications related to research projects in the world.

2 Architecture of an IVS system

This section will briefly describe the main components of an IVS system. Independent from the application, some IVS tasks are almost unavoidable, although their execution order and their mutual feedback can be designed in different ways, according to the architecture of the system and the peculiarities of the application itself. For instance, some systems first identify moving objects and then classify them to select only useful targets (e.g., vehicles or people); others, instead, provide target detection at the beginning and then they track selected objects only. Apart from some initial image enhancement and image processing steps, in any IVS system we can identify the following steps, even if their order can be different:

- *Foreground/object detection*: This task is devoted to detect what is of interest in the scene, in order to localize it and then track it along time. Typically, motion is an important cue and background is of no interest. For this reason, in stationary cameras this task is always provided by background suppression, where fixed, not moving areas are automatically separated from moving (foreground) areas. In more general scenarios with moving cameras or without a reference image, segmentation according with motion fields, color, texture, and so on have been adopted. Since the surveillance system detects and can be interested on vehicles, animals in addition to people, hereinafter the more general term of *moving object* is often used (see [10,11]).
- *Moving object tracking*: This task exploits time coherency to follow the same object along the time and the space. The dichotomy of tracking-by-detection or detection-by-tracking is still open [12]. When detection is easily provided in stationary camera, *appearance-based tracking* models working at pixel level are preferable (e.g., [13]). However, a complete survey on the topic was provided by Yilmaz *et al.* [14].
- *Object recognition*: This task consists of exploiting model coherency to provide object identification and classification, discerning between vehicles, people, bicycles, and so on, or even with a finer classification that can include the object identification (such as in the case of face recognition for identifying people). This problem was initially underestimated in surveillance since the type of objects is often predefined within the application (e.g., vehicles on road or people in an office), but it is mandatory in real scenarios, especially in the protection of critical infrastructures.
- *Scene understanding*: This high-level task is less defined than the previous ones and typically depend on the application. Some examples include the recognition of actions, events, and behaviors; the analysis of object trajectories in the scene; the estimation of crowd flows; and so on. This is typically the final step, and new generations of IVS systems must integrate a final step of high-level reasoning to assess the situation, infer possible dangerous or interesting events, and generate alarms.
- *Data fusion*: This optional task consists of fusing information coming from different sensor/camera sources to provide better detection and tracking, or to enlarge and enrich the recognition and understanding of the scene. As mentioned above, this task is becoming more popular due to the wide distribution of cameras and sensors for security purposes. In fact, the natural evolution of standard surveillance systems goes in the direction of enlarging the data availability using more cameras in parallel and consequently more processing modules for providing the previously discussed steps. In all these situations, data fusion is mandatory.

3 Examples of applications

3.1 LAICA project

Among the possible critical infrastructures, cities are the first that come in mind, given the high density of population and possible threat targets. This scenario, often referred to as *urban surveillance*, is one of the most challenging since the city is a complex entity with people and numerous sources of data/noise. The regional project LAICA (acronym for Laboratorio di Ambient Intelligence per una Città Amica—in English, Laboratory of Ambient Intelligence for a Friendly City) has been a 2-year (2005–2006) project funded by Regione Emilia-Romagna, Italy for a total budget of over 2 million euros and that involved universities, industries, and public administrations for a total of about 320 man-months.

The main objective of LAICA project is to explore the Ambient Intelligence (AmI) capabilities in a medium-sized Italian city such as Reggio Emilia. LAICA partners aim at defining innovative models and technologies for AmI in urban environments and at studying and developing advanced services for the citizens and the public officers in order to improve personal safety and prevent crimes. Multimedia and multimodal data have been collected from different sources, such as cameras, microphones, textual data about the traffic, the security, and the general situation of the city. The processed information has been made available to both police control centers and citizens by means of a dedicated Web site.

The foreseen services should be provided by a set of prototypal systems, as for instance, a system for the automatic monitoring of pedestrian subways by means of mobile and low-power audio and proximity sensors [15], a system that generates a feedback in pedestrian crossing systems to select the best duration of the green signal for the crossing [16], and a system for the automatic monitoring of public parks with a plethora of cameras (both fixed and PTZ) [17], also accounting for privacy issues [9].

The monitoring of public areas is more properly related to the topic of this chapter and requires the exploitation of multiple cameras for enhancing the security of people. Unfortunately, multiple cameras are useless if uncorrelated. The exploitation of the multiple viewpoints to correlate data from multiple cameras is often called *consistent labeling*, referring to the fact that the label/identity of moving objects is made consistent not only *over time* (as in the case of tracking from a single camera) but also *over space* (in the sense of different cameras). Often cameras' fields of view are disjoint, due to installation and cost constraints. In this case, the consistent labeling should be based on appearance only, basing the matching essentially on the color of the objects [18].

If the fields of view are overlapping, consistent labeling can exploit geometry-based computer vision, as we did in LAICA by proposing a novel method, called HECOL (Homography and Epipolar-Based COnsistent Labeling) [19]. The method takes into account both geometrical and shape features in a probabilistic framework. Homography and epipolar lines are computed to create relationships between cameras. When a new object is detected, the system checks the

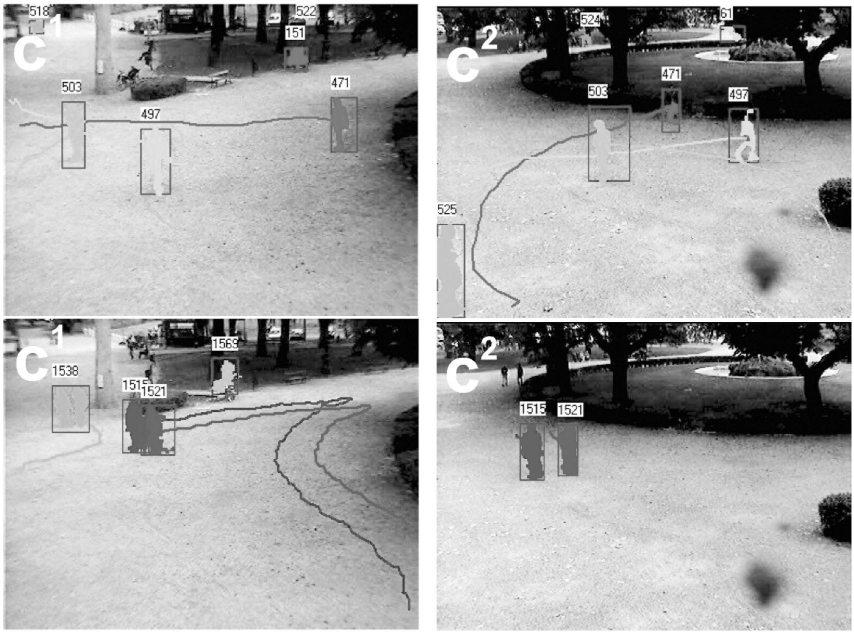

Figure 1: Examples of the HECOL system.

mutual correspondence of people using the axis of the objects precisely warped in the other field of view using epipolar lines. It accounts for the matching of the warped axis and the shapes of people. This makes the method particularly robust against segmentation errors and allows to disambiguate groups of people.

Figure 1 shows some examples of results of HECOL from a public park in Reggio Emilia.

3.2 THIS project

Transportation systems (fuel supply, railway network, airports, harbors, and inland shipping) are assets that are essential for the functioning of a society and economy and thus are critical infrastructures that can be effectively protected by an active surveillance system.

The European project THIS—Transport Hub Intelligent Video System (2010–2011)—deals with both new research approaches and concrete solutions for the safety and the security of transport hubs. THIS is carried on with the support of the Prevention, Preparedness and Consequence Management of Terrorism and Other Security-Related Risks Programme, European Commission, Directorate-General Justice, Freedom and Security.

In particular, the project THIS addresses automatic behavioral analysis through video processing, focused on crowded scenarios. THIS aims at a system performing human behavioral analysis, detaching what is usual from what is not, providing a reactive and hopefully proactive control task, preventing terroristic attacks or crime situations in public place. In order to learn what is normal or not, THIS proposes to use statistical inference enriched with contextual information. For example, "starting to run" in an exit zone could be abnormal and suspicious, but it becomes normal if the person is trying to reach a closing gate.

New computer vision technologies and software tools will be included in current available surveillance systems, equipped with web interfaces to be friendly managed by public operators. The final objective is to test research and emerging industrial solutions in real contexts to give a concrete answer to the possibility to prevent crime or terroristic attacks in public places, focusing the attention of the operators on meaningful situations in real time.

To evaluate the effectiveness of the developed system and to perform a preliminary requirement analysis, a number of use cases have been defined at the beginning of the project. These use cases are designed in order to serve three basic scenarios:

1. *Human interaction*, which deals with interactions among humans that need further investigation (e.g., people hugging and shaking hands, people quarrelling/fighting, etc.);
2. *Suspicious behavior*, which focuses on the detection of abnormal behaviors that could be indicators of potential criminal or terroristic incidents (e.g., individual running without an obvious reason, frequent visitor detection, abnormal trajectory analysis, etc.);
3. *Suspect surveillance*, which considers the continuous tracking of a suspected individual using the fixed cameras of a transport hub, which requires technologies such as consistent labeling [19], people reidentification [20], and so on.

The surveillance platform under development has been conceived to be as much flexible, customizable, and general as possible. Typical video surveillance systems consist at least of a single camera, connected with an embedded or general purpose computer with local or remote storage, display resources and related computer vision software. When more than one single camera processing is active at the same time, as usual in surveillance of critical infrastructure such as transport hubs, more knowledge about the scene can be obtained from the combination of the single outputs.

The overall surveillance tracking system can be seen as a set of clusters of overlapping cameras. Each node consists of a single camera processing, embedding the traditional single view stack of tasks. Inside each cluster, a strong interaction among nodes guarantees the consistent labeling by means of geometrical and appearance constraints. Information coming from each cluster are then merged and managed by a higher level processing [21]. The resulting framework is a three-layer architecture as depicted in the topmost part of Figure 2.

In addition to the tracking, a complete surveillance system should include several high-level modules to cope with the recognition and understanding points. Face detection, face recognition, action analysis, and event detection are some

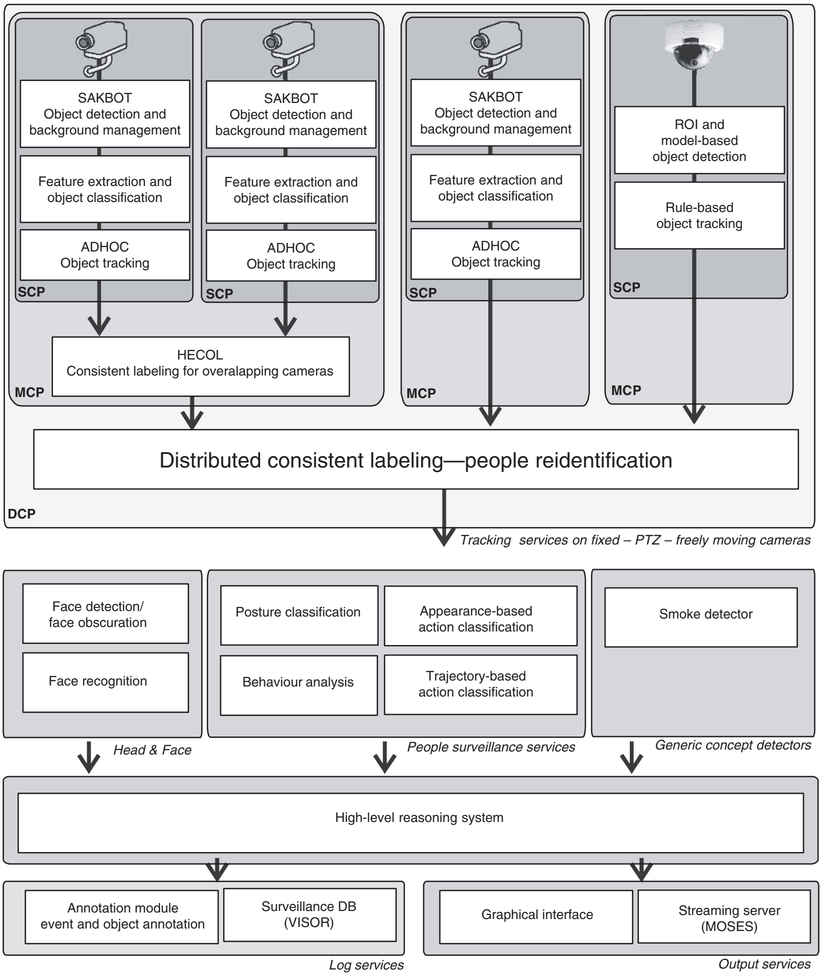

Figure 2: Multilayer and multiservice surveillance framework under development within the surveillance project THIS.

examples of functionalities of automatic people surveillance systems. Other modules can be added depending on the application and/or the research environment.

Some of these tasks should or could be performed on a subset of the installed cameras only or with a particular time schedule. Moreover, the same task (e.g., face detection) can be carried on using different techniques. A wide surveillance system should be able to take into account all these requirements. To this aim, THIS project proposes a Service-Oriented Architecture to model each system functionality, with an event-driven communication schema supporting this plethora of services. In Figure 2 a schema of the framework is reported, and in Table 1 the main services implemented in the prototype are described.

Table 1: List of the main surveillance services implemented in the framework.

People classifier	The HoG-based people classifier [22] is implemented as a service to detect people among the set of tracks, whenever they appear in the scene
Face detector	Two different face detectors are implemented in the framework: the well-known Viola & Jones of the OpenCV library and the face detection library by Kienzle *et al.* [23]
Posture classifier	The frame-by-frame posture of each person can be classified by means of the visual appearance. The implemented posture classification is based on projection histograms and selects the most likely posture among standing, sitting, crouching, and laying down [24]
Appearance-based action recognition	The action in progress is detected using features extracted from the appearance data. Walking, waving, and pointing are some examples of the considered actions. Two different approaches have been selected and implemented: The first is based on Hidden Markov Models, [25] and the second on action signature [26]
Trajectory-based action recognition	People trajectories (i.e., frame-by-frame positions of the monitored tracks) embed information about the people behavior; in particular, they can be used to detect abnormal paths that can be related to suspicious events. A trajectory classifier has been added in the system following the algorithm described in [27]
Smoke detector	The smoke detection algorithm proposed in [28] has been integrated in the system. The color properties of the object are analyzed accordingly to a smoke reference color model to detect if color changes in the scene are due to a natural variation or not. The input image is then divided in blocks of fixed sized and each block is evaluated separately. Finally a Bayesian approach detects whether a foreground object is smoke
Video streaming	MOSES (MObile Streaming for vidEo Surveillance) is a streaming system proposed by Gualdi *et al.* [29] that supports video streaming in different conditions, aiming at low-latency transmission over limited-bandwidth networks. In addition, the video stream is provided with a sufficient quality to be correctly analyzed by both human-based or computer-based video surveillance layers

3.3 Other examples

To complete the list of all the research projects or possible applications related to IVS is a paramount effort, which is out of the scope of this chapter. However, some of these are specifically related to the protection of critical infrastructures and will be here briefly introduced.

In the field of surveillance of cities and urban environments in general, besides the LAICA project already mentioned, it is worth mentioning the project UrbanEye[1] funded by the European commission within the Fifth Framework Programme (FP) and devoted to provide a comparative overview of institutional contexts of CCTV (Closed-Circuit TV) systems in Austria, Denmark, Germany, Hungary, Norway, Spain, and the United Kingdom by textual interpretation of policy documents, legislation, and the analysis of media talks in selected print media. The CAVIAR[2] project in Sixth FP also addressed the so-called city center surveillance aiming to provide semiautomatic analysis of videos for preventing nighttime crime and antisocial behavior problems, such as drunkenness, fights, vandalism, breaking and entering shop windows, and so on.

Also Seventh FP is really focused on the security research, and several projects have already been accepted. For instance, the VANAHEIM (Video/Audio Networked surveillance system enhAncement through Human-cEntered adaptIve Monitoring) project has started in February 2010 and will last 42 months. This project will address the study of innovative surveillance components for monitoring critical infrastructures such as underground transportation systems. Another example is the recent SUPPORT (Security UPgrade for PORTs) IP European project that started in June 2010 (and will last 48 months). SUPPORT is addressing potential threats on passenger life arising from intentional unlawful attacks on port facilities. The overall benefit will be the secure and efficient operation of European ports, enabling uninterrupted flows of cargo and passengers while suppressing attacks on high-value port facilities, illegal immigration, and trafficking of drugs, weapons, and illicit substances.

Also international communities, especially from United States, have demonstrated much interest in the use of IVS for surveillance of critical infrastructures. Among the many, one interesting example is the use of a multicamera IVS system for monitoring a public sport event, such as the SuperBowl XXXVII in 2003 in San Diego, CA. The research team of Prof. Mohan Trivedi at UCSD developed a system called DIVA (Distributed Interactive Video Arrays) [5] used to detect humans and animals in visually cluttered scenes on a 24-hour basis (using also a thermal camera) and to analyze the traffic flow on a road nearby stadium. Finally, an omnicamera in downtown San Diego has been used to simultaneously monitor traffic conditions and estimate the crowd size.

---

[1] http://www.urbaneye.net/index.html

[2] http://homepages.inf.ed.ac.uk/rbf/CAVIAR/caviar.htm

From the above list of research projects, it should be clear that the application of IVS systems for the protection of critical infrastructures is becoming more and more popular worldwide, thanks to the advances in the computer vision and pattern recognition technologies.

4 Conclusions

Surveillance of wide areas with several connected cameras integrated in the same automatic system is no more a chimera, but modular, scalable, and flexible architectures are mandatory to manage them. The explosion of requests of IVS systems for different scenarios leads research activity to explore many different dimensions in terms of both architectural issues and algorithms for scene recognition and interpretation.

Thus, many synergic fields spanning from hardware embedded systems, sensor networks, computer architecture on one side to image processing, computer vision, pattern recognition, and computer graphics on the other should be tightly integrated to cope with real-time surveillance applications and to provide reliable and effective answers to the security needs of critical infrastructures.

The most promising research directions in this field include the implementation of surveillance algorithms on embedded systems (to be applied on forests of small yet smart sensors), the higher level of understanding of the scene (to bridge the gap between the human operator and the IVS system), and the intelligent human–computer interfaces for forensic applications (which can guarantee an easy-to-use and fast retrieval of semantically meaningful information).

References

[1] R. Cucchiara & A. Prati, "Multicamera, distributed and heterogeneous sensor systems: Architectures and algorithms for people surveillance," *Proceedings of the First Workshop on Sensor- und Datenfusion—Architekturen und Algorithmen*, Berlin, Germany, 2009.

[2] G. Bocchetti, F. Flammini, C. Pragliola, & A. Pappalardo, "Dependable integrated surveillance systems for the physical security of metro railways," *Proceedings of the Third ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC 2009)*, Como, Italy, 2009, pp. 1–7.

[3] S. Fleck & W. Strasser, "Towards secure and privacy sensitive surveillance," *Proceedings of ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC 2010)*, Atlanta, GA, USA, 2010, pp. 126–132.

[4] R. Cucchiara, "Multimedia surveillance systems," *Proceedings of the Third ACM International Workshop on Video Surveillance and Sensor Networks (VSSN 2005)*, Singapore, 2005, pp. 3–10.

[5] M.M. Trivedi, T.L. Gandhi, & K.S. Huang, "Distributed interactive video arrays for event capture and enhanced situational awareness," *IEEE Intelligent Systems*, vol. 20, 2005, pp. 58–66.

[6] Y.-li Tian, L. Brown, A. Hampapur, M. Lu, A. Senior, & C.-fe Shu, "IBM smart surveillance system (S3): Event-based video surveillance system with an open and extensible framework," *Machine Vision and Applications*, vol. 19, 2008, pp. 315–327.

[7] N. Haering, P. Venetianer, & A. Lipton, "The evolution of video surveillance: An overview," *Machine Vision and Applications*, vol. 19, 2008, pp. 279–290.

[8] R.T. Collins, A.J. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y. Tsin, D. Tolliver, N. Enomoto, O. Hasegawa, P. Burt, & L. Wixson, "A system for video surveillance and monitoring," *Technical Report of Camegie Melion University (CMU)*, CMU-RI-TR-00-12, 2000.

[9] R. Cucchiara, A. Prati, & R. Vezzani, "A system for automatic face obscuration for privacy purposes," *Pattern Recognition Letters*, vol. 27, 2006, pp. 1809–1815.

[10] M. Piccardi, "Background subtraction techniques: A review," *Proceedings of IEEE SMC 2004 International Conference on Systems, Man and Cybernetics*, The Hague, The Netherlands, 2004, pp. 3099–3104.

[11] R.J. Radke, S. Andra, O. Al-Kofahi, & B. Roysam, "Image change detection algorithms: A systematic survey," *IEEE Transactions on Image Processing*, vol. 14, 2005, pp. 294–307.

[12] M. Andriluka, S. Roth, and B. Schiele, "People-tracking-by-detection and people-detection-by-tracking," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2008)*, Anchorage, Alaska, USA, 2008, pp. 1–8.

[13] T. Zhao & R. Nevatia, "Tracking multiple humans in complex situations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, 2004, pp. 1208–1221.

[14] A. Yilmaz, O. Javed, & M. Shah, "Object tracking: A survey," *ACM Computing Surveys*, vol. 38, 2006, p. 13.

[15] P. Zappi, E. Farella, & L. Benini, "A PIR-based wireless sensor node prototype for surveillance applications," *Proceedings of European Workshop on Wireless Sensor Networks (EWSN 06)*, Zurich, Switzerland, 2006, pp. 26–27.

[16] A. Broggi, R.L. Fedriga, & A. Tagliati, "Pedestrian detection on a moving vehicle: An investigation about near infrared images," *Intelligent Vehicles Symposium, 2006 IEEE*, Meguro-ku, Japan, 2006, pp. 431–436.

[17] S. Calderara, R. Cucchiara, & A. Prati, "Group detection at camera handoff for collecting people appearance in multi-camera systems," *Proceedings of the IEEE International Conference on Video and Signal Based Surveillance (AVSS'06)*, Sydney, NSW, Australia, 2006, p. 36.

[18] R. Vezzani, D. Baltieri, & R. Cucchiara, "Pathnodes integration of standalone particle filters for people tracking on distributed surveillance systems," *Image Analysis and Processing—ICIAP 2009*, P. Foggia, C. Sansone, and M. Vento, eds, Springer Berlin/Heidelberg, Germany, 2009, pp. 404–413.

[19] S. Calderara, A. Prati, & R. Cucchiara, "HECOL: Homography and epipolar-based consistent labeling for outdoor park surveillance," *Computer Vision and Image Understanding*, vol. 111, 2008, pp. 21–42.

[20] M. Farenzena, L. Bazzani, A. Perina, V. Murino, & M. Cristani, "Person re-identification by symmetry-driven accumulation of local features," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, San Francisco, CA, USA, 2010, pp. 2360–2367.

[21] R. Vezzani & R. Cucchiara, "Event-driven software architecture for multi-camera and distributed surveillance research systems," *Proceedings of the First IEEE Workshop on Camera Networks—CVPRW*, San Francisco, CA, 2010, pp. 1–8.

[22] N. Dalal & B. Triggs, "Histograms of oriented gradients for human detection," *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR05)*, IEEE Computer Society, Washington, DC, 2005, pp. 886–893.
[23] W. Kienzle, G. Bakir, M. Franz, & B. Scholkopf, "Face detection—efficient and rank deficient," *Advances in Neural Information Processing Systems*, vol. 17, 2005, pp. 673–680.
[24] R. Cucchiara, C. Grana, A. Prati, & R. Vezzani, "Probabilistic posture classification for human-behavior analysis," *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, vol. 35, 2005, pp. 42–54.
[25] R. Vezzani, M. Piccardi, & R. Cucchiara, "An efficient Bayesian framework for on-line action recognition," *Proceedings of the IEEE International Conference on Image Processing*, Cairo, Egypt, 2009.
[26] S. Calderara, R. Cucchiara, & A. Prati, "Action signature: A novel holistic representation for action recognition," *IEEE*, Santa Fe, NM, 2008.
[27] S. Calderara, A. Prati, & R. Cucchiara, "Learning people trajectories using semi-directional statistics," *Proceedings of Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, Genova, Italy, 2009.
[28] P. Piccinini, S. Calderara, & R. Cucchiara, "Reliable smoke detection system in the domains of image energy and color," *6th International Conference on Computer Vision Systems, Vision for Cognitive Systems*, Santorini, Greece, 2008.
[29] G. Gualdi, A. Prati, & R. Cucchiara, "Video streaming for mobile video surveillance," *IEEE Transactions on Multimedia*, Santorini, Greece, vol. 10, 2008, pp. 1142–1154.