Scanning the Issue/Technology

Special Issue on Video Communications, Processing, and Understanding for Third Generation Surveillance Systems

I. Introduction

A surveillance system can be defined as a technological tool that assists humans by providing an extended perception and reasoning capability about situations of interest that occur in the monitored environments. Human perception as well as reasoning are constrained by the capabilities and limits of human senses and mind to simultaneously collect, process and store limited amount of data. For example:

- only information coming from a limited spatial area can be directly sensed and processed by the human at a given time;
- the complexity of the situations that can be analyzed is usually limited to events, occurring at different time instants, that can be associated by reasoning with their common causes.

Surveillance systems provided varied degrees of assistance to humans evolved in an incremental way according to the progress in surveillance technologies [1]. We will describe in the following sections the details of the successive generations of surveillance systems that increasingly utilize a larger set of sensors as well as more flexible and robust processing strategies.

This Special Issue focuses on the problems of last generation surveillance systems and highlights solutions to these problems that are based on a stronger integration of techniques for multisensor data acquisition, communications and processing. This integration is possible by the common "full digital" perspective on which the techniques used by new systems are based. Next generation surveillance systems can be considered as an emerging application field requiring multidisciplinary expertise going from signal and image processing, to communications and computer vision. This multidisciplinary view is common to many applications in the information and communications technology (ICT) domain, such as videoconferencing, ambient intelligence, etc. There is a growing interest in surveillance applications due to the growing availability of cheap sensors and processors at reasonable costs. There is also a growing need from the public for improved safety and security in large urban environments and improved usage of resources of public infrastructure. This, in conjunction with the increasing maturity of algorithms and techniques, is making possible the application of this technology in various application sectors such as security, transportation, and the automotive industry. In particular, the problem of remote surveillance of unattended environments has received growing attention in the last years, especially in the context of:

a) safety in transport applications [2], [3], such as monitoring of railway stations [4], [5], underground stations [6], [7], airports [8]–[10] and airplane routes [11]–[13], motorways [14], [15], urban and city roads [16]–[23], maritime environments [24]–[27];
b) safety or quality control in industrial applications, such as monitoring of nuclear plants [28] or industrial processing cycles [1]–[3];
c) improved Security for people lives, such as monitoring of indoor or outdoor environments like banks [29], supermarkets [6], car parking areas [30], waiting rooms [31], buildings [32], [33], etc., remote monitoring of the status of a patient [34], remote surveillance of the human activity [35]–[47];
d) military applications for surveillance of strategic infrastructures [48], [49], enemy movements in the battlefield [50], [51], air surveillance [52], [53].

In order to satisfy a market potentially so large, strong research innovations are required that allow surveillance engineers and end-users to take advantage of innovative communication solutions, processing, and understanding methods that are developed by researchers. The goal of this Special Issue is to point out the key aspects and technological trends of the last generation of surveillance systems.

While several modalities of sensing such as audio, video, and chemical sensors are useful in monitoring; we chose to concentrate on those applications where visual information plays the most important role. Video communication, processing, and understanding can be considered as a fundamental modality for surveillance applications.

This is due to several factors.

- Temporally organized visual information is the major human source of information about the surrounding environment.

Publisher Item Identifier S 0018-9219(01)08438-9.

- As the number of cameras increase, event monitoring by personnel is rather boring, tedious, and error- prone. The automatic preprocessing of the video information by a surveillance system can act as a prefilter to human validation of the events. Thus, it is a natural mechanism to manage the complexity of monitoring a large site. In addition, a high-level interface presenting the events in a site is a most user-friendly and widely acceptable presentation.
- The cost of the video sensor is considerably lower compared to other sensors when one takes into account the area of coverage and event analysis functionality provided by using video as the sensing modality for monitoring.
- A large body of knowledge exists in the areas of robust and fast digital communication, video processing, and pattern recognition. These facilitate the development of effective and robust real-time systems.
- Digital video presents stringent throughput requirements for a multimedia communication system in terms of robustness and real-time performance.

Nevertheless, video information can be acquired, processed, and transmitted in different ways, and we have provided a panoramic view of such modalities in this issue.

Video communications aspects are fundamental in surveillance systems [54]–[59]. Data are acquired by distributed sources and then are usually transmitted to some remote control center. An important communication requirement is the bandwidth that should be lower for the down-link (from the control center to the sensors) than for the up-link (from the sensors to control center). Another important aspect is the security of the transmission. In many applications, surveillance data must be transmitted over open networks with multiuser access characteristics [18]. Information protection on such networks is a critical issue for maintaining privacy in the surveillance service. On the other hand, paternity of surveillance data can be very important for effective use for law enforcement purposes. Therefore, legal requirements necessitate the development of watermarking and data-hiding techniques for secure sensor identity assessment. Video processing and understanding requirements in surveillance systems are more severe than in classical computer vision systems due to the high variability and irregularity of the monitored scenes. Such variability has several consequences in required processing tools. From one point of view, it makes it necessary to use more sophisticated image processing algorithms for signal preprocessing and filtering. On the other hand, highly variable scene conditions imply the necessity of selecting robust scene description and pattern recognition methods. The automatic capability to learn and adapt to changing scene conditions and the learning of statistical models of normal event patterns are emerging issues in surveillance systems [42], [60]. The learning system provides a mechanism to flag potentially anomalous events by the discovery of the normal patterns of activity and flagging the least probable ones. Two major constraints that impact the deployment of these systems in the real world include real-time performance and low cost [61]. Moreover, the multisensor aspect of a surveillance system constitutes a rather important direction for improving algorithms [2]. Multisensor systems can take advantage from processing either the same type of information acquired from different spatial locations or information acquired by sensors of different type (e.g., video cameras, microphones, etc.) on the same monitored area [3]. Appropriate processing techniques and new sensors providing the real-time information related to different scene characteristics can help both to enlarge the size of monitored environments and to improve performances of alarm detection in areas monitored by more sensors.

II. Review of the State of the Art

Electronic video surveillance systems that have been proposed in literature can be classified under a technological perspective as belonging to three successive generations. The three generations follow the evolution of communications, processing, and storage and they have evolved in recent years with the same increasing speed of such technologies. Obviously, different categorizations can be established (see, e.g., [1]) that are based on different aspects of surveillance: for example, categories have been proposed to classify surveillance systems according to the degree of awareness of observed people being monitored. An excellent historical perspective is presented in [1] of the basic scientific discoveries that allowed surveillance video devices, storage media, and image transmission techniques to be progressively developed. Early breakthroughs in optics, including the discovery of lenses and concepts leading to the pinhole camera model, are shown to be as important as the more recent event understanding and recording tools (Daguerre [63]). The capability of observing and recording images from distant places has been originally oriented to monitor what happens in heaven. However, more prosaic observation of what happens on earth has been discovered by video-based surveillance to be as interesting; however, surveillance of events occurring on Earth poses ethical problems as such events often involve humans and the right to monitor can be in conflict with the individual privacy rights of the monitored people. These privacy problems largely depend on the shared acceptance of the surveillance task as a necessity by the public at large with respect to a given application. Another technological breakthrough fundamental to the development of surveillance systems is the capability of remotely transmitting and reproducing images and video information [e.g., TV broadcasting and the successive use of video signal transmission and display in close circuit TV systems (CCTV)]. CCTVs operative on the market and providing data at acceptable quality can be found dating back to 1960. The availability of CCTVs can be considered as the starting point that allowed on-line surveillance to be possible, and 1960 can be considered the starting date of the first generation surveillance systems.

First video surveillance systems (**1GSS**) (1960–80) basically extend human perception capabilities in a spatial sense. More "eyes" (i.e., video cameras) are used to display

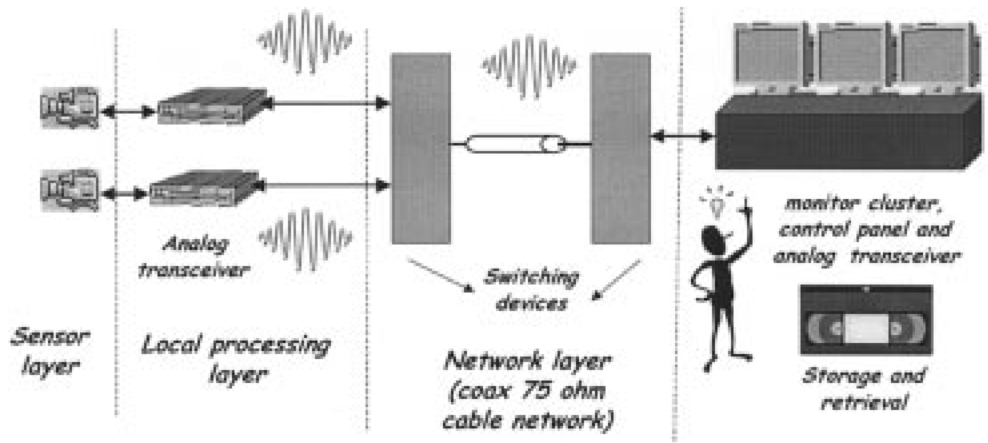

Fig. 1 Architectural example of first generation video-based surveillance system (1960–1980).

analog visual signals from multiple remote locations in a single physical location (i.e., the control room). 1GSSs are based on analog signal and image transmission and processing (Fig. 1). In these systems, video data from a set of cameras viewing remote scenes (sensor layer) are presented to the human operators after analog communication (local processing layer) of the video signal. Human operators analyzed video streams through a large set of monitors, where the scenes monitored by multiple cameras were multiplexed and presented in a periodic and predefined order. An added value of 1GSS is given by the acquired capability of telepresence of a human with respect to a remote place in a certain instant. Some major drawbacks of these systems have to do with the reasonably small attention span of operators that may result in a high miss rate of the events of interest. From a communications point of view, these systems suffered from the main problems of analog video communications: i.e., high bandwidth requirements, poor allocation flexibility, etc. Storage of video surveillance tapes remained a problem until the mid-1970s, when analog storage on VHS and similar media alleviated this problem.

The main limitations of the first generation systems are due to the following points strictly related to analog processing and transmission level.

— A large bandwidth is usually required that limits the number of sensors to be used [57].
— Analog video is subject to noise in transmission and the stored information suffers from degradations in image quality during playback [54]–[59].
— On-line alarm detection for a large set of monitored sites is difficult as they are related to visual inspection of monitors by human operators with limited attention spans [64].
— Off-line archival and retrieval of information on significant events of interest is difficult due to the large amount of tapes to be stored and reexamined.

It is clear from the above points that if either the spatial extent of the area being monitored or the complexity of events increases, then the only practical solution for real-time event detection using 1GSS is to increase the number of operators, i.e., to increase the number of parallel human processors for signals associated with events.

Starting from 1980, rapid improvements in the different basic technologies emerged: the improved resolution of video cameras and the availability of low-cost computers are two basic breakthroughs that facilitated intense research on algorithms for video processing and detection of events. In parallel, communications improvements during the 1980s led to CCTVs with improved robustness at reduced costs. In this technological evolution, *second generation surveillance systems* (**2GSS**) (1980–2000) correspond to the maturity phase of analog 1GSSs; they benefited from early advances in digital video communications (e.g., digital compression, bandwidth reduction, and robust transmission) and processing methods that provide assistance to the human operator by prescreening of important visual events. Some of these systems have been studied since the late 1980s until now in the context of different international research programs [65], [66] and have carried to prototypical products showing the feasibility of digital, intelligent attention focusing systems on video from limited sets of cameras.

In particular, 2GSS research addressed many areas with increased results in real-time analysis and segmentation of two-dimensional (2-D) image sequences [67], identification and tracking of multiple objects in complex scenes [68]–[73], human behavior understanding [35]–[45], multisensor data-fusion [74], intelligent man–machine interfaces [75]–[77], performance evaluation of video processing algorithms [78], [79], wireless and wired broad-band access networks [80]–[83], new signal processing for video compression, and multimedia transmission for video-based surveillance systems [84]–[90], etc.

Most research efforts during the period of 2GSSs have been spent on the development of automated real-time event detection techniques for video surveillance. As we have mentioned before, the availability of automated methods would greatly facilitate the monitoring of large sites with numerous cameras as the automated event detection step allows for prefiltering and presentation of the relevant events.

In this way, the augmented perception capability in 2GSSs allows for a significant increase in the amount of simultaneously monitored data and, in addition, provides alarm data directly relevant to the cognitive monitoring tasks. Humans

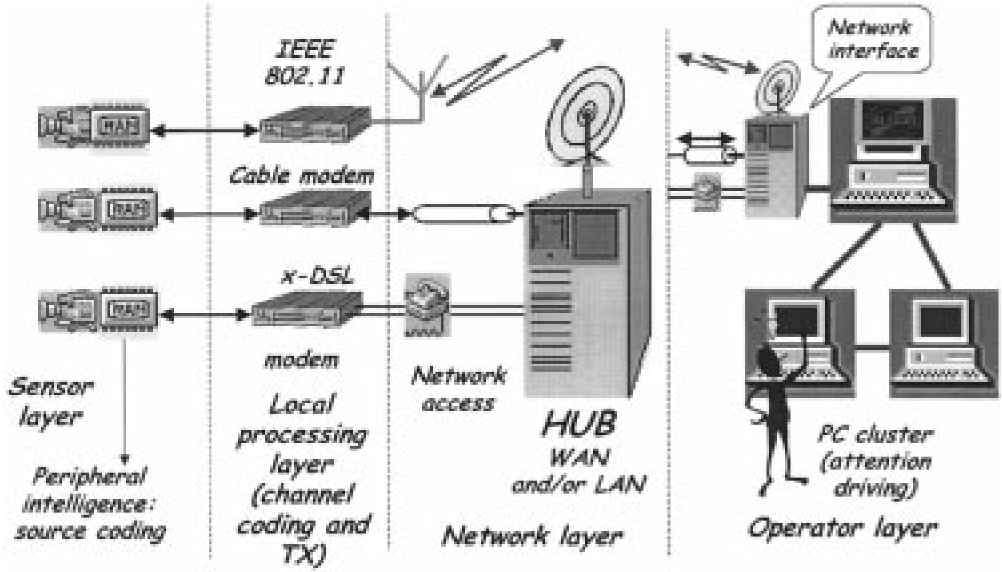

Fig. 2 Architectural example of third generation video-based surveillance system.

and animals are provided this ability through the use of preattentive mechanisms. It has been shown that there is evidence that neural nets implementing motion detection are used by the brain to capture human attention on specific sections of the human retina. Simple examples of multisensor extensions of this phenomenon are provided by the capability of humans to focus their sensors toward spatial areas from which specific sounds have been heard. However, 2GSSs have been able to only provide solutions with intermediate levels of digital video signal transmission and processing [80]–[90], i.e., they occasionally include digital methods in system subparts to solve local and isolated problems.

The main goal of *third generation surveillance systems* (**3GSS**) is to provide "full digital" solutions to the design of surveillance systems, starting at the sensor level, up to the presentation of mixed symbolic and visual information to the operators (see Fig. 2). In this sense, they take advantage of progress in low cost, high performance computing networks and in the availability of digital communications on heterogeneous, mobile, and fixed broad-band networks [56], [57].

In Fig. 2, an example of 3GSS is presented where video cameras constitute the sensor layer, while the peripheral intelligence and the transmission devices form the local processing layer. Sensor and local processing layers can be physically organized together in a so-called *intelligent camera*. The local processing layer uses digital compression methods to save bandwidth resources. The principal component of the network layer is the intelligent hub: the main functionality of the intelligent hub is the application-oriented fusion of data coming from lower-level layers. At the operator layer, an active interface is presented to the operator. This interface assists the operator by focusing his/her attention to a subset of interesting events. Communications are entirely in a digital form. The communication medium could be fixed wireless LANs or mobile digital devices (e.g., GPRS digital mobile phones) as well as broad-band media such as optical fibers, coax cables, or twisted pairs.

Research work on distributed real-time video processing techniques on intelligent, open, and dedicated networks is expected to provide more and more interesting results. This will be largely due to the availability of increased computational power at reasonable costs, advanced video processing/understanding methods, and multi-sensor data fusion. At the same time, a 3GSS can take advantage from the evolution of multimedia digital broadband communications in both wireless and wired domains. In particular, progress in the design of high-bandwidth access networks makes it possible to forecast widespread use of these systems by residential users for different applications. However, these surveillance systems would present specific requirements that necessitate the dedicated research and development of new tools.

This Special Issue is aimed at providing a global view of research efforts that are driving the development of 3GSSs as well as to provide an insight into the industrial perspectives of research centers developing them.

III. Technical Challenges of Deploying Third Generation Systems

We have seen that the main objective of full digital 3GSSs is to facilitate the efficient data communication, management, and extraction of events in real-time video from a large collection of sensors. To achieve this goal, improvements in automatic recognition functionalities and digital multiuser communications strategies are needed. Technology meeting the requirements for the recognition algorithms includes computational speed, memory usage, remote data access, multiuser communications between distributed processors, etc. The availability of this technology greatly facilitates 3GSS development and deployment.

From the point of view of augmentation of human perception and monitoring capabilities in 3GSSs, the 3GSS alleviates the human from monitoring a collection of video monitors and, in addition, would assist the human in tasks that are rather cumbersome (i.e., fall outside normal human spatial and temporal cognitive abilities) to do with traditional systems. For instance, real-time person tracking in a crowded scene is a tough task for a human to perform with a single video displayed on the monitor. Another improvement of 3GSSs is that online tools can be built to assist humans with event management.

A. Technological Viewpoint

If we consider technological aspects, one of the major technological basis of 3GSSs is the availability of robust, high-bandwidth digital multimedia transmissions over wide-band channels. Another technological basis has been the availability of embedded digital sensors that directly process locally acquired digital data. As progress in 3GSSs' intelligent sensors are being made, we are seeing the deployment of hubs capable of performing limited local digital video processing functions based on embedded DSPs. In addition, there is an increase in the amount of computing power per unit cost for use in the central control rooms and intelligent hubs, thus allowing automated intelligent processing to be done at the control center or in intermediate surveillance stations. Therefore, the driving technological push in 3GSSs is based on three main aspects.

- Wide-band digital communications and surveillance networking.
- Rapid decrease in processing hardware cost.
- Appearance of embedded intelligence subsystems (sensors and hubs).

Thanks to the availability of more evolved and powerful communications, sensors, and processing units, the architectural choice in 3GSSs can potentially become highly variable and flexibly customized to obtain a desired performance level. Therefore, the system architecture starts to represent a key issue; for example, the different level of distribution of intelligence can lead preattentive detection methods either closer to the sensors or distributed at different levels in a computational processing hierarchy. Another source of variability is due to the use of heterogeneous networks (wireless or wired) and transmission modalities both in terms of source and channel coding and in terms of multiuser access techniques. Spatial and temporal coding scalability can be very useful for reducing the amount of information to be transmitted by each camera depending on the intelligence level of the camera itself, while multiple access techniques are a basic tool to allow a large number of sensors to share a communication channel in the most efficient and robust way. Surveillance network management techniques are also necessary in 3GSSs to coordinate distributed intelligence modules in order to obtain a optimal performances as well as to adapt system behavior depending on the variety of conditions occurring either in a scene or in systems' parameters. All these tools are critical to design efficient systems. For example, the number of cameras supported by a system can vary to a large degree depending on both the level of intelligence embedded in each camera and on the channel capacity available for messages sent by cameras. Finally, a further evolution is the integration among surveillance networks based on sensors of either different types such as audio, radar or always visual but oriented toward completely different functionalities (e.g., face detection, fingerprinting) and sensor types (e.g., standard perspective cameras or catadioptric sensors, i.e., sensors with mirrors).

The major technological improvements expected in 3GSSs can be structured onto different generality levels. This depends on the major complexity of these systems with respect to previous generations. Moreover, we can suppose that, due to such complexity, the development of a 3GSS system with all the characteristics underlined in the following cannot be reached until ten years from now. This also opens the problem of identifying successive progressive steps inside 3GSSs that can reasonably be integrated at successive stages into a single system.

Let us first analyze major improvements expected at different levels.

At a general level, a 3GSS should support:

- multiple services related to different users accessing to the same set of data acquired by a surveillance network (controlled multiuser accessibility);
- flexible changing of the functionalities assigned either to a single cell or to a group of cells depending on the active services as well as on operating conditions (cell reconfigurability).

A surveillance service should be complete and it should allow data accessibility both for direct alarm generation and for off-line inspection, i.e., it must include:

- a user oriented, sufficiently extended number of functionalities associated with a number of sensorial cells sufficient to provide a spatial surveillance support appropriate for the task (completeness);
- an alarm generation mechanism satisfying real-time alarm generation user requirements (real-time response);
- Distributed digital memorization capabilities and local databases accessible for a given time from the event and covering a sufficiently extended period (off-line recording).

Each functionality should be characterized by measurability, robustness, efficiency, multimodal sensor support, and adaptability with respect to both processing and communications. In particular, each functionality should be associated with the following.

- A computational model of a detection method appropriate to identify events of interest from available signal representations (computability).
- A measurable performance metric depending on the operating conditions (measurability).
- A performance behavior that should degrade gracefully with respect to the presence of various environmental conditions; such conditions should include the possibility for a functionality that can be applied to recorded, compressed data, by considering compression rate as an external condition (robustness).

Table 1
Real-World Applications

Application Domain	*Primary Benefits*	*Intelligent Functionality desired*	*Cost and Performance requirements*
Public area monitoring Large facility monitoring	Safety, Security	Person/vehicle detection, tracking and event analysis	Low system cost, false alarm/detection requirements rather stringent
Building exterior and interior monitoring, Parking Garage monitoring	Security, Safety, Access Control, Building Automation	Person/vehicle detection, parking space monitoring, license plate recognition, face recognition	High-end market High reliability desired in access control. Illumination is controlled / unconstrained
Subway, Highway, Tunnel monitoring, Transportation applications	Safety, Security, Resource management and Improved quality of service	People detection and tracking, vehicle, truck detection/tracking, classification of type of objects, recognition of events	Few high-end systems exist in the market, Very low false alarm rates. All weather and illumination conditions
Indoor monitoring (Malls, lobbies, Banks shopping complexes)	Security and Safety	Person detection, tracking, event analysis	Low cost systems, minimal false alarms

— A modifiable processing behavior to detect events of interest depending on environmental scene conditions (processing adaptability).

— A modifiable communication strategy depending on conditions of the channels (communications adaptability).

— A high ratio between performances with respect to employed computational and bandwidth resources (efficiency).

— An appropriate selection of sensors organized into system cells in order to provide data necessary to detect events of interest (multimodal sensorial support).

IV. Research Impact on Real-World Product Development

We have seen how the technological trends impact the research and development of the 3GSSs. The design, development, and deployment of these systems in the real world are influenced by a variety of factors including: the availability of sophisticated algorithms, the integration of the algorithms into the system form, and the validation that the system designed meets end user requirements. The industrial trends in CCTV systems are to incorporate intelligent processing functionality into these systems. High-end systems are being offered that take advantage of the broad-band communication capabilities and the intelligent algorithms available. However, their acceptance in the real world has been rather slow, mainly due to prohibitive cost of these systems and due to the end-user acceptance of these products (for a good discussion on end-user concerns and a market analysis of the security industry please see the paper from Pavlidis *et al.* of Honeywell Research in this Special Issue). Early use of CCTV systems has been in large public installations (i.e., subway systems, large public areas) for improved safety and security, in military installations, private buildings, banks, and in shopping centers. More increasingly, video monitoring systems are being used in medium-scale shopping centers and in small shops. These are still based on 2GSS technology. Evaluation of a 2GSS system is primarily based on the quality of the image or video being presented to the user, on the number of video streams that can be monitored effectively. However, in a 3GSS system this is not the case. The intelligence functionality in a 3GSS system introduces the fundamental issue of validation of the intelligence component to verify that the alarm generation software meets user requirements. Since the end-users do not understand computer vision or signal processing technologies, their expectations for this technology are rather high at first glance. It is not uncommon for a highway authority official to expect people/vehicle detection and tracking error rates of less than 1% in all weather conditions, a task that is rather daunting even for humans. There is a need for 3GSS system researchers and designers to understand realistic use case scenarios of these systems and to translate end-user requirements to design practical and efficient systems. In Table I, we categorize real-world applications, their functional requirements, and cost/performance requirements.

The major application areas for 3GSSs are in the area of public monitoring. This is necessitated by rapid growth of metropolitan localities and by the growing need to provide improved safety and security to the general public. Other factors that drive the deployment of these systems include effective resource management, providing rapid emergency assistance, etc. The market for security and surveillance systems is slated to grow from about $650 billion in the current year to about $1.225 billion in the year 2006 worldwide.

Some factors that currently impede the deployment of these systems include:

1) system costs for given performance;
2) robustness of system functions with respect to complexity of input video (e.g., outdoor/natural illumination conditions, all weather conditions);
3) lack of standards for quantification of performance of these systems;
4) high costs, tediousness of tests and validation;
5) high-level vision functions providing semantics in video are rather error prone and generate too many false positives;
6) automated systems need to provide self-diagnosis when a scenario that is not modeled is encountered.

The system costs are rather prohibitive currently if one examines the level of performance required. Video surveillance is a visual task that is boring yet easy for a human operator to perform during short attention spans. End-users cannot comprehend the difficulty in the automation of such visual tasks. Due to their lack of understanding of vision systems, unrealistic performance requirements are often set. Nevertheless, the false alarm rates per camera for an event detection task should be rather low. This is driven by the psychological need for the human operator to trust the automated system. If the automated system generated too many false alerts, the human would tend to ignore the automated system and, hence, the intelligence function will be turned off. The problem is compounded when many types of events are automatically generated. The false alarms just add up. Typical system requirements for a people detection task in highways, for instance, is close to 100% detection with near zero false alarms per day under all weather conditions. A false alarm in this case is the detection of a change in the scene as a person. Another system requirement is the reaction time, i.e., the time it takes the system for an alarm to be generated, for these systems. Typical reaction times may vary depending on the event, but it is reasonable to expect reaction times of the order of a few seconds.

Another major stumbling block in incorporating these intelligence functions in real-world systems is the lack of robustness, the inability to test and validate these systems under variety of usage cases, and the lack of quantification of performance of these systems. A major requirement in automated systems is the ability to self-diagnose when the video data is not usable for analysis purposes. For instance, when CCD cameras are used in an outdoor highway application, it is often the case that during certain times of the day there is direct lighting of the camera lens from sunlight; a situation that renders the video useless for monitoring purposes. Another example of such a scenario is a weather condition such as heavy snowfall during which the contrast levels are such that people detection at a distance is rather difficult to do. Thus, in these scenarios, it is useful to have a system diagnostic that alerts the end-user of the unavailability of the automated intelligence functions. Ideally, the function that evaluates the unavailability of a given system should estimate whether the input data is such that the system performance can be guaranteed to meet given user-defined specifications. In addition, the system should gracefully degrade in performance as the complexity of data increases. This is a very open research issue that is crucial to the deployment of these systems.

Performance evaluation of these systems, therefore, is a major open research issue. There is now a dedicated IEEE workshop on performance evaluation of tracking systems (PETS) that attempts to bring researchers to evaluate algorithms on common datasets to identify the algorithms strengths/limitations. However, there is a lack of realistic datasets and industrial input in these forums. Video databases that facilitate the systematic evaluation of the performance of various intelligent processing functions are needed. These databases should capture essentially all the variability in the scene conditions (e.g., day, night, day to night transitions, all object types, event types, dry, rainy, snow, foggy conditions) to effectively determine the situations under which the algorithms are effective and meet requirements. There is a need for performance metrics and well-agreed definitions for evaluating system components and the total system performance. Product development will benefit for the systematic comparisons of available methods. Testing and validation of these systems is rather costly and tedious due to the manual labor involved in validation. Intelligence functions can be built to have enough logged information to validate the alarms generated, while a periodic sampling/logging of the video data along with manual examination by a person is necessary to identify potentially missed alarms.

The first functionalities that we will see in the 3GSSs are intelligent detection and tracking functions with limited event analysis capabilities. Research systems currently have demonstrated these functionalities; see, for instance, [91]. The complexity of these event analysis methods is still rather low. They are primarily algorithms evaluating trajectories of movement patterns of people/vehicles to identify potential anomalies. The algorithms operate mainly in light pedestrian traffic conditions. More complicated event analysis functions will be needed to deal with moderate flow conditions. These will require multiple object tracking, reasoning, and interpretation of events.

V. Special Issue Contents

In the previous sections, some of the main aspects were highlighted related to the current state of the art, technology, and industrial applications trends with respect to video surveillance systems. This Special Issue aims at providing a deeper insight in this topic by providing to the readers a balanced list of contributions of academic and industrial research aspects in communications, processing and understanding. As the reader will see from the papers of the Special Issue and as one can expect from the real-world problems explained in the previous section, main problems currently considered are related with real-time either distributed or centralized processing and robustness issues in multisensor surveillance networks. We hope that the invited papers presented by some of the more active research groups in this field will provide at the same time a sufficiently extended framework of current research status and new ideas for people who are interested in contributing to this interesting field where academic approaches and industrial viewpoint can successfully meet to provide solutions from which real-world end-users can benefit. Nevertheless, we are also sure that this Special Issue necessarily covers only a limited part of the global work carried on in this field by not directly describing research of other academic and industrial groups in the world. Therefore, we invite interested readers to go through the references in various papers and in other Special Issues published in books and specialistic journals (e.g., [1]–[3], [61], [80], [90]–[92]) to enlarge the view we provided in this issue.

Referring to the contents of this special issue, we now present an overview of each of the invited and peer-reviewed published papers.

A. Change Detection and Background Extraction by Linear Algebra

(Invited Paper)

Durucan and Ebrahimi

The first paper in the Special Issue deals with a key issue in surveillance systems, i.e., optimal approaches to reduce the cardinality of data to be considered by further processing steps to obtain real-time scene descriptors. Change detection and background evaluation is particularly important in scenes observed by fixed cameras and can be managed in different ways depending by scene characteristics. In this first paper on change detection techniques as applied to video surveillance, the authors present an overview of several methods and discuss an innovative method that they have successfully applied in prototypical surveillance systems. The method is based on a physical luminance model and uses algebraic considerations to derive an estimation of the area of interest of an image with respect to an estimated background.

B. Into the Woods: Visual Surveillance of Noncooperative and Camouflaged Targets in Complex Outdoor Settings

(Invited Paper)

Boult, Micheals, Gao, and Eckmann

This paper discusses the current state of the art in video-based target detection with particular attention to the problem of surveillance and tracking of noncooperative and camouflaged targets in cluttered outdoor settings. Since for these domains, the detection phase is crucial, the authors discuss mainly techniques for change detection. Then, they present an innovative approach, called quasi-connected components (QCC), for performing spatio-temporal grouping. QCC combines gap filling, thresholding-with-hysteresis, and spatio-temporal region merging. The last part of the paper briefly review the tracking component of the system as well as the target geo-location, network communication, and user interface. Finally, the authors discuss the performance evaluation of the system, as measured by an external evaluation group.

C. Image Authentication Techniques for Surveillance Applications

(Invited Paper)

Bartolini, Tefas, Barni, and Pitas

The problem of image authentication in digital video surveillance systems is considered in this paper by authors coming from two European universities very active in the watermarking research field. In particular, this paper provides an introductory overview to watermarking techniques where different approaches are discussed with their relative merits as compared to the considered application. This paper introduces the interesting viewpoint of designing watermarking algorithms in systems where quality is assessed not on the basis of a subjective/objective visual judgment but on the basis of indirect results i.e., automatic system decisions, like event detection in surveillance systems.

D. Distributed Architectures and Logical-Task Decomposition in Multimedia Surveillance Systems

(Invited Paper)

Marcenaro, Oberti, Foresti, and Regazzoni

Third generation video surveillance systems use distributed intelligence functionality. An important design issue is to decide the granularity at which the tasks can be distributed based on available computational resources, network bandwidth, and task requirements. The paper investigates the impact of distributed processing and communication techniques on the design of 3GSSs. The authors illustrate how the distribution of intelligence can be achieved by dynamic partition of all the logical processing tasks, including event recognition and communication. The dynamic task allocation problem is studied through the use of a computational complexity model for representation and communication tasks. The computational power of the intelligent cameras and the channel capacity of the bandwidth transmission are shown to be important parameters that affect the performance of the total system.

E. Multiple Camera Tracking of Interacting and Occluded Human Motion

(Invited Paper)

Dockstader and Tekalp

This paper describes a multicamera system for tracking interacting human motion based on multiple layers of temporal filtering coupled by a Bayesian belief network. The system uses a distributed platform, where a dedicated processor is applied to process each independent video stream representing a distinct view of some scene, to achieve real-time performance and to reduce overcome problems with occlusions and articulated motion. Each image of the monocular sequence is processed to extract interesting 2-D features (i.e., a set of image points to be tracked) of human motion. These measurements are used together with an estimate of the 3-D state vector representing the velocity and position of features in a 3-D Cartesian space as the input of a predictor-corrector filter that produces an estimate of the 2-D state vector. 2-D state vectors coming from each view of the system provide a vector of independent observations for a Bayesian belief network which fuses them to compute the most likely vector of 3-D state estimates given the available data. To maintain temporal continuity, the network is followed with a layer of Kalman filters that updates the 3-D state estimates. Experiments on a home environment with several people in motion demonstrate the superiority of the proposed approach in tracking accuracy with respect to data fusion methods based on averaging.

F. Algorithms for Cooperative Multisensor Surveillance

(Invited Paper)

Collins, Lipton, Fuiyoshi, and Kanade

The Robotic Institute at Carnegie Mellon University (CMU) has created a Video Surveillance and Monitoring Lab. The team working in this laboratory has developed several video-understanding algorithms to perform cooperative, multisensor surveillance. An overview of these algorithms, integrated into a multicamera system, is described in this paper. The proposed system uses a distributed network of active video sensors to monitor activities in a cluttered outdoor environment. Video understanding algorithms are used to automatically detect people and vehicles, to localize and track them into a geo-spatial reference system and to classify them. Results from each single camera system are integrated into a coherent overview of the dynamic scene by multi-sensor fusion algorithms running on a central control room. Results are shown to a remote operator through a graphical user interface that provides the user with 2-D and 3-D synthetic views of the environment. Detected objects are displayed as dynamic agents.

The feasibility of the real-time functioning of the surveillance system has been demonstrated within a multicamera test-bed system developed on the CMU campus.

G. Urban Surveillance Systems: From the Laboratory to the Commercial World

(Invited Paper)

Pavlidis, Morellas, Tsiamyrtzis, and Harp

This paper describes a system developed in an industrial research center for the monitoring of a large building site parking lot with distributed set of cameras. The paper offers an industrial perspective to the security market as well as the end-user concerns. It discusses a system "DETER" that is used to detect and track pedestrians and vehicles in a parking lot using a distributed set of sensors. Various aspects of the system including background adaptation, object detection and tracking, trajectory analysis for threat identification, and visualization of the results from various sensors are presented. In addition, a qualitative as well as quantitative evaluation of the system performance is presented.

H. Design, Analysis, and Engineering of Video Monitoring Systems: An Approach and a Case Study

(Invited Paper)

Greifenhagen, Comaniciu, Niemann, and Ramesh

The problem of including a quantitative statistical performance evaluation model in the design of an industrial oriented multisensor surveillance system is the problem considered in this paper. The authors show first a general methodology by which a surveillance problem can be divided in a set of submodules, each characterized by a precise statistical input–output relation. They show how performance of complex chains of such modules can be predicted in a statistical sense on the basis of probabilistic knowledge on input data. The industrial value of the paper is given by the case of study shown, where the problem of integrating data coming from an omni-directional camera to obtain an estimate of people position in a indoor scene can be used to point the optical axis of a different mobile camera toward the face of the observed people. The used performance evaluation model used in the design phase allows one to evaluate pointing error depending on the position of the observed people in the field of view of the omni-directional camera and to fix accordingly intrinsic and extrinsic parameters of the mobile camera to optimize the view of the observed face.

I. Aerial Video Surveillance and Exploitation

(Invited Paper)

Kumar, Sawhney, Samarasekera, Hsu, Tao, Guo, Hanna, Pope, Wildes, Hirvonen, Hansen, and Burt

This paper from an industrial research center describes a state-of-the-art aerial video surveillance system developed over several years of efforts for the Department of Defense Advanced Projects Agency in the U.S. The paper describes a framework for aerial video surveillance using video cameras. Aerial video surveillance is done delineating the video into components corresponding to static scene geometry, the dynamic objects, and the appearance of static/dynamic objects in the scene. The delineation is done based on 2-D/3-D alignment of dynamic imagery. Models that are progressively increasing complexity are invoked to delineate the static and dynamic components of the scene and efficiently represented for exploitation in various surveillance tasks. The paper discusses key components of the framework, including frame-to-frame alignment and the extraction of motion layers, mosaicing of static background components to form panaromas, independent tracking of moving objects, extraction of the geo-location of the video and tracked objects, and enhanced visualization of the video by reprojection and merging of the video with reference imagery and/or digital terrain maps. The system produces meta-data along with the video that allows one to perform aerial mapping, dynamic scene visualization over time, temporal change detection, etc.

CARLO S. REGAZZONI, *Guest Editor*
University of Genoa
Genoa I-16145, Italy

VISVANATHAN RAMESH, *Guest Editor*
Siemens Corporate Research Inc.
Princeton, NJ 08540 USA

GIAN LUCA FORESTI, *Guest Editor*
University of Udine
Udine 33100, Italy

ACKNOWLEDGMENT

The Guest Editors wish first to thank M. Kunt and the Proceedings Board for precious suggestions and generous encouragement to develop a framework proposal where to fit the original idea of this Special Issue. We wish also to thank J. Calder for his continuous and cooperative support during the period of preparation of this issue. Warmest thanks go to the invited authors who enthusiastically accepted our invitation of preparing high quality papers for this issue as well as to the following experts whose precious voluntary contribution as reviewers made it possible improvement of the quality of invited papers: T. Ellis, A. Venetsanopuolos, G. Thieil, O. Silven, Y. Kuno, T. Kalker, L. Cinque, F. Ramos, E. Memin, M. Tekalp, I. Pitas, T. Boult, D. Comaniciu, J. Llinas, V. Morellas, I. Pavlidis, I. Haritaoglu, J. Illingworth, T. Ebrahimi, N. Paragios, R. Kumar, D. Aubert, V. Roberto, J. Ferryman, F. Roli, M. Mustafa, P. Remagnino, S. Santini, G Jones.

Moreover, the Guest Editors wish to thank F. Oberti and L. Marcenaro for their assistance in editorial activities.

REFERENCES

[1] J. K. Petersen, *Understanding Surveillance Technologies*. Boca Raton, FL: CRC Press, 2001.

[2] C. S. Regazzoni, G. Fabri, and G. Vernazza, *Advanced Video-Based Surveillance Systems*. Norwell, MA : Kluwer, 1998.

[3] G. L. Foresti, P. Mahonen, and C. S. Regazzoni, *Multimedia Video-Based Surveillance Systems: Requirements, Issues and Solutions*. Norwell, MA : Kluwer, 2000.

[4] H. Susama and M. Ukay, "Application of image processing for railways," *Q. RTRI*, vol. 30, no. 2, pp. 74–81, 1989.

[5] U. Urfer, "Integration of systems and services in central monitoring stations (CMS)," in *Proc. IEEE Int. Carnahan Conf. Security Technology*, 1995, pp. 343–350.

[6] M. Bogaert, N. Chelq, P. Cornez, C. S. Regazzoni, A. Teschioni, and M. Thonnat, "The PASSWORD project," in *Proc. Int. Conf. Image Processing*, Chicago , IL, 1996, pp. 675–678.

[7] C. S. Regazzoni and A. Tesei, "Distributed data fusion for real-time crowding estimation," *Signal Process.*, vol. 53, pp. 47–63, 1996.

[8] E. F. Lyon, "The application of automatic surface lights to improve airport safety," *IEEE AES Syst. Mag.*, pp. 14–20, 1993.

[9] M. Braasch, M. DiBenedetto, S. Braasch, and R. Thomas, "LAAS operations in support of airport surface movement, guidance, control and surveillance: Initial test results," in *IEEE Proc. Position Location and Navigation Symp.*, 2000, pp. 82–89.

[10] G. Galati, M. Ferri, P. Mariano, and F. Marti, "Advanced integrated architecture for airport ground movements surveillance," in *Radar Conf.*, 1995, pp. 282–287.

[11] H. Wang, T. Kirubarajan, and Y. Bar-Shalom, "Precision large scale air traffic surveillance using IMM/assignment estimators," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 35, pp. 255–266, Jan. 1999.

[12] B. Sridhar and G. B. Chatterji, "Computationally efficient conflict detection methods for air traffic management," in *Proc. American Control Conf.*, vol. 2, 1997, pp. 1126–1130.

[13] G. Donohue, "Vision on aviation surveillance systems," in *Proc. IEEE Int. Radar Conf.*, 1995, pp. 1–4.

[14] G. L. Foresti and B. Pani, "Monitoring motorway infrastructures for detection of dangerous events," in *IEEE Proc. Int. Conf. Image Analysis and Processing*, Venice, Italy, 1999, pp. 1144–1147.

[15] J. M. Manendez, L. Salgado, E. Rendon, and N. Garcia, "Motorway surveillance through stereo computer vision," in *IEEE Proc. 33rd Annu. Int. Carnahan Conf. Security Technology*, 1999, pp. 197–202.

[16] D. Koller, K. Daniilidis, and H. Nagel, "Model-Based object tracking in monocular image sequences of road traffic scenes," *Int. J. Comput. Vis.*, vol. 10, pp. 257–281, 1993.

[17] J. Malik, D. Koller, and J. Weber, "Robust multiple car tracking with occlusion reasoning," in *Eur. Conf. Computer Vision*, Stockolm, Sweden, 1994, pp. 189–196.

[18] S. H. Park, K. Jung, J. K. Hea, and H. J. Kim, "Vision-Based traffic surveillance system on the internet," in *Proc. 3rd Int. Conf. Computational Intelligence and Multimedia Applications (ICCIMA '99)*, 1999, pp. 201–205.

[19] J. E. Boyd, J. Meloche, and Y. Vardi, "Statistical tracking in video traffic surveillance," in *Proc. 7th IEEE Int. Conf. Computer Vision*, vol. 1, 1999, pp. 163–168.

[20] A. F. Toal and H. Buxton, "Spatio-temporal reasoning within a traffic surveillance system," in *Proc. 2nd Eur. Conf. Computer Vision*, S. Margherita , Ed., Italy, 1992, pp. 884–892.

[21] J. M. Blosseville, "Image processing for traffic management," in *Advanced Video-Based Surveillance Systems*, C. S. Regazzoni, G. Vernazza, and G. Fabri, Eds. Norwell, MA: Kluwer, 1999, pp. 67–75.

[22] C. P. K. Sherwood, "Traffic surveillance and control systems for the area of Hong Kong," in *9th Int. Conf. Road Transport Information and Control*, Apr. 21–23, 1998, pp. 191–194.

[23] S. A. Hamid, S. A. Rahman, and J. J. Steed, "The introduction of traffic surveillance and control on the privatised express ways in Malaysia," in *9th Int. Conf. Road Transport Information and Control*, Apr. 21–23, 1998, pp. 200–206.

[24] R. B. Olsen, P. Bugden, Y. Andrade, P. Hoyt, M. Lewis, H. Edel, and C. Bjerkelund, "Operational use of RADARSAT SAR for marine monitoring and surveillance," in *Int. Symp. Geoscience and Remote Sensing, IGARSS '95*, vol. 1, 1995, pp. 224–226.

[25] A. N. Ince and E. Topuz, "The design and computer simulation of a maritime surveillance system," in *Int. Conf. Radar, Radar 97*, 1997, pp. 653–656.

[26] K. Takasaki, T. Sugimura, and S. Tanaka, "Comparison of sea traffics in Tokyo and Osaka Bays with JERS-1/OPS data," in *Proc. Int. Geoscience and Remote Sensing Symp., IGARSS '96*, pp. 79–81.

[27] J. G. Sanderson, M. K. Teal, and T. J. Ellis, "Target identification in complex maritime scenes," in *6th Int. Conf. Image Processing and its Applications*, vol. 2, 1997, pp. 463–467.

[28] C. A. Rodriguez, J. A. Howell, H. O. Menlove, C. M. Brislawn, J. N. Bradley, P. Chare, and T. Gorten, "NUCLEAR video image processing for nuclear safeguards," in *Proc. IEEE 29th Annu. Int. Carnahan Conf. Security Technology*, 1995, pp. 355–363.

[29] B. B. Berson, R. S. Wallance, and E. L. Schwartz, "A miniaturized active vision system," in *Proc. 2nd Int. Conf. Pattern Recognition*, vol. 4, 1992, pp. 58–61.

[30] M. J. Cattle, "The use of digital CCTV in an airport car-park application," in *Proc. IEEE 29th Annu. Int. Carnahan Conf. Security Technology*, 1995, pp. 180–185.

[31] E. Stringa and C. S. Regazzoni, "Content-based retrieval and real time detection from video sequences acquired by surveillance systems," in *IEEE Int. Conf. Image Processing*, vol. 3, Chicago, IL, Oct. 4–7, 1998, pp. 138–142.

[32] C. Lin and R. Nevatia, "Building detection and description from a single intensity image," *Comput. Vis. Image Understand.*, vol. 72, no. 2, pp. 101–121, 1998.

[33] L. Vergara and P. Bernabeu, "Automatic signal detection applied to fire control by infrared digital signal processing," *Signal Process.*, vol. 80, no. 4, pp. 659–669, 2000.

[34] W. Millesi, M. J. Truppe, F. Watzinger, A. Wagner, and R. Ewers, "Image guided surgery extended by remote stereotactic visualization," in *Lecture Notes in Computer Science*, J. Troccaz, E. Grimson, and R. Mosges, Eds. Berlin, Germany: Springer, 1997, vol. 1205, pp. 813–821.

[35] I. Haritaoglu, D. Harwood, and L. S. Davis, "W4: Real-time surveillance of people and their activities," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, pp. 809–830, Aug. 2000.

[36] J. Heikkila and O. Silven, "A real-time system for monitoring of cyclists and pedestrians," in *Proc. 2nd IEEE Workshop on Visual Surveillance (VS'99)*, 1999, pp. 74–81.

[37] S. Hongeng, F. Bremond, and R. Nevatia, "Representation and optimal recognition of human activities," in *Proc. Int. Conf. Computer Vision and Pattern Recognition (CVPR2000)*, 2000, pp. 818–825.

[38] D. M. Gavrila, "The analysis of human movement and its application for visual surveillance," in *Proc. 2nd IEEE Workshop on Visual Surveillance (VS'99)*, 1999.

[39] N. M. Oliver, B. Rosario, and A. P. Pentland, "A Bayesian computer vision system for modeling human interactions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 831–843, 2000.

[40] Q. Cai and J. K. Aggarwal, "Tracking human motion using multiple cameras," in *Proc. 13th Int. Conf. Pattern Recognition*, Vienna, Austria, Aug. 25–29, 1996, pp. 68–72.

[41] Y. Ricquebourg and P. Bouthemy, "Real-time tracking of moving persons by exploring spatio-temporal image slices," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 797–808, 2000.
[42] A. Galata, N. Johnson, and D. Hogg, "Learning behavior models of human activities," in *Proc. British Machine Vision Conf.*, U.K., 1999.
[43] J. Aranda, J. Amat, and M. Fragola, "A multitracking system for trajectory analysis of people in a restricted area," in *Proc. 4th Int. Workshop on Time-Varying Image Processing and Moving Object Reconigtion*, Florence, Italy, 1993.
[44] J. A. Freer, B. J. Beggs, H. L. Fernandez-Canque, F. Chevrier, and A. Goryashko, "Automatic intruder detection incorporating intelligent scene monitoring with video surveillance," in *Proc. Eur. Conf. Security and Detection, ECOS 97*, 1997, pp. 109–113.
[45] F. Bremond and M. Thonnat, "Tracking multiple nonrigid objects in video sequences," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, pp. 585–591, 1998.
[46] H. Buxton and S. Gong, "Visual surveillance in a dynamic and uncertain world," *Artif. Intell.*, vol. 78, no. 1-2, pp. 431–459, 1995.
[47] C. Benabdelkader, P. Burlina, and L. Davis, "Single camera multiplexing for multi-target tracking," in *Multimedia Videobased Surveillance Systems. Requirements, Issues and Solution*, C. S. Regazzoni, G. L. Foresti, and P. Mahonen, Eds. Norwell, MA: Kluwer, 2000, pp. 130–142.
[48] D. A. Pritchard, "System overview and applications of a panoramic imaging perimeter sensor," in *Proc. IEEE 29th Annu. Int. Carnahan Conf. Security Technology*, 1995, pp. 420–425.
[49] U. Oppelt, "New possibilities for video applications in the security field," in *Proc. IEEE 29th Annu. Int. Carnahan Conf. Security Technology*, 1995, pp. 426–435.
[50] B. Peters, J. Meehan, D. Miller, and D. Moore, "Sensor link protocol: Linking sensor systems to the digital battlefield," in *Proc. IEEE Military Communications Conf.*, vol. 3, 1998, pp. 919–923.
[51] M. T. Fennell and R. P. Wishner, "Battlefield awareness via synergistic SAR and MTI exploitation," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 13, pp. 39–43, Feb. 1998.
[52] J. S. Draper, S. Perlman, C. K. Chuang, M. Hanson, L. Lillard, B. Hibbeln, and D. Sene, "Tracking and identification of distant missiles by remote sounding," in *Proc. IEEE Aerospace Conf.*, vol. 4, 1999, pp. 333–341.
[53] G. A. V. Sickle, "Aircraft self reports for military air surveillance," in *Proc. IEEE Digital Avionics Systems Conf.*, vol. 2, 1999, pp. 2–8.
[54] C. S. Regazzoni, C. Sacchi, and C. Dambra, "Remote cable-based video surveillance applications: the AVS-RIO project," in *Proc. ICIAP99*, Venice , Italy, Sept. 27–29, 1999, pp. 1214–1215.
[55] F. Soldatini, P. Mahonen, M. Saaranen, and C. S. Regazzoni, "Network management within an architecture for distributed hierarchical digital surveillance systems," in *Multimedia Videobased Surveillance Systems. Requirements, Issues and Solutions*, C. Regazzoni, G. Foresti, and P. Mahonen, Eds. Norwell, MA: Kluwer, pp. 143–157.
[56] P. Mahonen and M. Saaranen, "Broadband multimedia transmission for surveillance applications," in *Multimedia Videobased Surveillance Systems. Requirements, Issues and Solutions*, C. S. Regazzoni, G. L. Foresti, and P. Mahonen, Eds. Norwell, MA: Kluwer, pp. 173–185.
[57] K. Pahlavan and A. H. Levesque, "Wireless data communications," *Proc. IEEE*, vol. 82, no. 9, pp. 1398–1430, 1994.
[58] ——, *Wireless Information Networks*. New York: Wiley, 1995.
[59] S. Glisic and B. Vucetic, *Spread Spectrum CDMA Systems for Wireless Communications*. Norwood, MA: Artech House, 1997.
[60] M. Walter, A. Psarrou, and S. Gong, "Learning prior knowledge and observation augmented density models for human behavior recognition," in *Proc. British Machine Vision Conf.*, U.K., 1999.
[61] G. L. Foresti and C. S. Regazzoni, "Video processing and communications in real-time surveillance systems," *J. Real-Time Imaging*, vol. 7, no. 3, 2001.
[62] G. Gernsheim, H. Helmut, A. Allison, and L. J. M. Daguerre, *The History of the Diorama and the Daguerreotype*. New York: Dover, 1968.
[63] L. J. M. Daguerre, "Histoire et description des procedes du daguerreotype et du diorama Daguerre's Manual,", 1839.
[64] C. H. M. Donold, "Assessing the human vigilance capacity of control room operators," in *Proc. Int. Conf. Humans Interfaces in Control Rooms, Cockpits and Command Centres*, 1999, pp. 7–11.
[65] ESPRIT Program, European Union. [Online]. Available: http://www.newcastle.research.ec.org/esp-syn/all-ac-index.html
[66] VSAM Program, USA . [Online]. Available: http://www.cs.cmu.edu
[67] S. M. Smith, "ASSET-2: Real-time motion segmentation and object tracking," *Real Time Imaging*, vol. 4, pp. 21–40, 1998.
[68] G. L. Foresti, "Object detection and tracking in time-varying and badly illuminated outdoor environments," *Opt. Eng.*, vol. 37, no. 9, pp. 2550–2564, 1998.
[69] Z. Li and H. Wang, "Real-time 3-D motion tracking with known geometric models," *Real Time Imaging*, vol. 5, pp. 167–187, 1999.
[70] P. J. L. V. Beek, A. M. Tekalp, N. Zhuang, I. Celasun, and M. Xia, "Hierarchical 2-D mesh representation, tracking and compression for object-based video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, pp. 617–634, 1999.
[71] D. B. Gennery, "Visual tracking of known 3D objects," *Int. J. Comput. Vis.*, vol. 7, no. 3, pp. 243–270, 1992.
[72] D. G. Lowe, "Robust model-based motion tracking through the integration of searching and estimation," *Int. J. Comput. Vis.*, vol. 8, no. 2, pp. 113–122, 1992.
[73] F. Meyer and P. Bouthemy, "Region-based tracking using affine motion models in long image sequences," *Computer Vision, Graphics and Image Processing: Image Understanding*, vol. 60, no. 2, pp. 119–140, 1994.
[74] P. K. Varshney, "Multisensor data fusion," *Electron. Commun. Eng. J.*, vol. 9, no. 6, pp. 245–253, 1997.
[75] A. Pentland, "Looking at people: Sensing for ubiquitous and wearable computing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 1, pp. 107–119, 2000.
[76] D. S. Faulus and R. T. Ng, "An expressive language and interface for image querying," *Mach. Vis. Applicat.*, vol. 10, no. 2, pp. 74–85, 1997.
[77] D. P. Haanpaa and G. P. Roston, "An advanced haptic system for improving man-machine interfaces," *Comput. Graphics*, vol. 21, no. 4, pp. 443–449, 1997.
[78] F. Oberti, E. Stringa, and G. Vernazza, "Performance evaluation criterion for characterizing video surveillance systems," *Real-Time Imaging J.*, vol. 7, no. 3, 2001.
[79] T. Kanungo, M. Y. Jaisimha, J. Palmer, and R. M. Haralick, "A methodology for quantitative performance evaluation of detection algorithms," *IEEE Trans. Image Processing*, vol. 4, no. 12, pp. 1667–1673, 1995.
[80] W. W. Lu, "Special issue on multidimensional broadband wireless technologies and services," *IEEE Trans. Commun.*, vol. 89, Jan. 2001.
[81] K. Fazel, P. Robertson, O. Klank, and F. Vanselow, "Concept of a wireless indoor video communications system," *Signal Process. Image Commun.*, vol. 12, no. 2, pp. 193–208, 1998.
[82] P. Scotton, "Compression and transmission of video signals over high speed networks with rate based congestion control," *Signal Process.*, vol. 36, no. 3, pp. 392–392, 1994.
[83] P. Batra and S. F. Chang, "Effective algorithms for video transmission over wireless channels," *Signal Process. Image Commun.*, vol. 12, no. 2, pp. 147–166, 1998.
[84] B. S. Manjunath, T. Huang, A. M. Tekalp, and H. J. Zhang, "Introduction to the special issue on image and video processing for digital libraries," *IEEE Trans. Image Processing*, vol. 9, no. 1, pp. 1–2, 2000.
[85] E. Stringa and C. S. Regazzoni, "Real-time video-shot detection for scene surveillance applications," *IEEE Trans. Image Processing*, vol. 9, no. 1, pp. 69–79, 2000.
[86] G. Bjontegaard, K. O. Lillevold, and R. Danielsen, "A comparison of different coding formats for digital coding of video using MPEG-2," *IEEE Trans. Image Processing*, vol. 5, no. 8, pp. 1271–1276, 1996.
[87] H. Cheng and X. Li, "Partial encryption of compressed images and videos," *IEEE Trans. Signal Processing*, vol. 48, no. 8, pp. 2439–2451, 2000.
[88] J. Benoispineau, F. Morier, D. Barba, and H. Sanson, "Hierarchical segmentation of video sequences for content manipulation and adaptive coding," *Signal Processing*, vol. 66, no. 2, pp. 181–201, 1998.
[89] N. Vasconcelos and A. Lippman, "Statistical models of video structure for content analysis and characterization," *IEEE Trans. Image Processing*, vol. 9, no. 1, pp. 3–19, 2000.
[90] T. Ebrahimi and P. Salembier, "Special issue on video sequence segmentation for content-based processing and manipulation," *Signal Processing*, vol. 66, no. 2, pp. 123–124, 1998.
[91] R. Collins, A. Lipton, and T. K. Kanade, "Special issue on video surveillance and monitoring," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 745–746, 2000.
[92] S. Maybank and T. Tan, "Introduction—Surveillance," *Int. J. Comput. Vis.*, vol. 37, no. 2, pp. 173–173, 2000.

Carlo Regazzoni (Senior Member, IEEE) was born in Savona, Italy, in 1963. He received the Laurea degree in electronic engineering and the Ph.D. degree in telecommunications and signal processing from the University of Genoa, Italy, in 1987 and 1992, respectively.

Since 1998, he has been Professor of Telecommunication Systems in the Engineering Faculty of the University of Genova. Since 1998, he has been responsible for the Signal Processing and Telecommunications (SP&T) Research Group at the Department of Biophysical and Electronic Engineering (DIBE), University of Genova, that he joined in 1987. His main current research interests are multimedia and nonlinear signal and video processing, signal processing for telecommunications, multimedia broad-band wireless, and wired telecommunications systems. He has been involved in research on multimedia surveillance systems since 1988. He has been co-organizer and chairman of the first two International Workshops on Advanced Video Based Surveillance, held in Genova, Italy, 1998 and Kingston, U.K., 2001. He has also organized several Special Sessions in the same field at International Conferences [Image Analysis and Processing, Venice 1999 (ICIAP99), European Signal Processing Conf. (Eusipco2000), Tampere Finland, 2000]. He has been responsible for several EU research and development projects dealing with video surveillance methodologies and applications in the transport field (ESPRIT Dimus, Athena, Passwords, AVS-PV, AVS-RIO). He has been also responsible for several research contracts with Italian industries; he served as a referee for international journals and as reviewer for EU in different research programs. He is a consultant for the EU Commission for the definition of the 6th research framework program in the Ambient Intelligence domain. He is co-editor of the books *Advanced Video-based Surveillance Systems* (1999) and *Multimedia Surveillance Systems* (2000). He is author or co-author of 43 papers in international scientific journals and of more than 130 papers presented at refereed international conferences.

Dr. Regazzoni is a Member of AEI and IAPR

Visvanathan Ramesh received the B.S. degree in engineering from the College of Engineering, Guindy, Madras, India in 1984, the M.S. degree in electrical engineering from Virginia Tech in 1987, and the doctoral degree from the University of Washington, where he defended his Ph.D dissertation titled "Performance characterization of image understanding algorithms" in 1994.

He is currently a Senior Member of Technical Staff and Project Manager of the real-time imaging effort in the Imaging Department at Siemens Corporate Research, Princeton, NJ. At Siemens, he has focused on the research and development of statistical methods for real-time video analysis functions such as object detection, tracking and action recognition. His most recent work is focused on the research of how contextual models and domain knowledge influence the selection of algorithms and its parameters for a given video analysis application. He supervises Ph.D students in several universities, including CMU, UIUC, Lehigh and U of Rochester. He also has advised several M.S. students from various universities in Europe. He has been actively involved in image and video understanding research in low- and mid-level vision over the past 12 years and has published numerous publications on the topic. His primary objective is to build robust image and video analysis systems and to quantify robustness of IU algorithms. During the course of his Ph.D. work, he developed a systems engineering methodology for computer vision algorithm performance characterization and design. He has also focused on the development of software environments for computer vision. Besides his deep involvement in the Unix version of GIPSY (a general image processing system), he was a member of the ARPA Image Understanding Environment Committee, the committee that designed the IUE (an object oriented environment for Image Understanding Research). He was also part of a team that helped design the Java Advanced Imaging Specification (Sun Microsystem's Java API for advanced imaging). He has published several papers in the computer vision field, with large emphasis in the area of performance analysis of vision systems. He is a co-author of a paper on real-time tracking that got the best paper award in CVPR 2000. His broad research interests are pattern recognition, computer vision, artificial intelligence, and biomedical engineering.

Gian Luca Foresti (Senior Member, IEEE) was born in Savona, Italy, in 1965. He received the Laurea degree (*cum laude*) in electronic engineering and the Ph.D. degree in computer science from the University of Genoa, Italy, in 1990 and in 1994, respectively.

In 1994, he was visiting Professor at University of Trento, Italy. Since 1998, he has been Professor of the Computer Science at the Department of Mathematics and Computer Science (DIMI), University of Udine. His main interests involve artificial neural networks, multisensor data fusion, computer vision and image processing, and multimedia databases. Techniques proposed found applications in the following fields: automatic video-based systems for surveillance and monitoring of outdoor environments (e.g., underground stations, railway lines, motorway, etc.), vision systems for autonomous vehicle driving and/or road traffic control, 3-D scene interpretation, and reconstruction. He is author or co-author of more than 100 papers published in International Journals and Refereed International Conferences. He was general co-chair, chairman and member of Technical Committees at several conferences. He has been co-organizer of three Special Sessions on video-based surveillance systems at International Conferences (ISATA97, ISATA98, ICIAP99). He has contributed to five books in his area of interest, and he is co-author of the book *Multimedia Systems for Visual-Based Surveillance* (Kluwer, 2000).

Dr. Foresti was Guest Editor of the Special Issue of *Real Time Imaging Journal* on "Video Processing and Communications in Real Time Video-Based Surveillance" and recently he was Guest Editor of a Special Issue of the PROCEEDINGS OF THE IEEE on "Video Communications, Processing and Understanding for Third Generation Surveillance Systems." He was an invited speaker at the NATO School on Multisensor Data Fusion, at Pitlocry, U.K., July 2000. He has served as a reviewer for several international journals and for the European Union in different research programs (MAST III, Long Term Research, Brite-CRAFT). He has been responsible for DIMI for several European and national research projects in the field of video-based surveillance for unattended outdoor environments. In February 2000, he was appointed as an Italian Member of the Information Systems Technology (IST) panel of the NATO-RTO. He is a Senior Member of IAPR.