Intelligent Video Systems and Analytics: A Survey

Honghai Liu, *Senior Member, IEEE*, Shengyong Chen, *Senior Member, IEEE*, and Naoyuki Kubota, *Member, IEEE*

***Abstract*—Recent technology and market trends have demanded the significant need for feasible solutions to video/camera systems and analytics. This paper provides a comprehensive account on theory and application of intelligent video systems and analytics. It highlights the video system architectures, tasks, and related analytic methods. It clearly demonstrates that the importance of the role that intelligent video systems and analytics play can be found in a variety of domains such as transportation and surveillance. Research directions are outlined with a focus on what is essential to achieve the goals of intelligent video systems and analytics.**

***Index Terms*—Behavior detection, computer vision, intelligent video system (IVS), surveillance, video analytics, visual context recognition.**

I. Introduction

INTELLIGENT video systems (IVS) and intelligent video analytics (IVA) have been substantially growing from practical needs in the past decade, being driven by a wide range of applications in transportation and healthcare [1]. In the last years, the video analytics market has been about 60% annual compound growth.[1] By traditional video surveillance, a watchstander often faces the duty of staring at hundreds of screens. In fact, it is a crucial challenge if a person has to monitor everything effectively even between only two screens. Human eyes get tired quickly after few minutes if one has to draw its attention among different screens. Even with just a single screen for a period as long as 30 min, one can miss more than 80% of the activity in the scene.[2]

Regarding the terminology, Elliott recently defines that an IVS is "any video surveillance solution that utilizes technology to automatically, without human intervention, process, manipulate and/or perform actions to or because of either the live or stored video images" [2]. IVS embeds computer vision technologies into video devices such as cameras, encoders, routers, digital video recorders (DVRs), network video recorders (NVRs), and other video management and storage devices [3]. An IVS provides a constant, unblinking eye on any scene. With its IVA, the system helps government, public and commercial organizations to transform video surveillance into a real-time, proactive, event-driven process. The virtual operator stays focused by introducing a level of automation to video monitoring. Alert conditions may be set with real-time processing algorithms to deliver information for a security team to react swiftly and take action. Alerts may also be sent via wireless, Internet, and telephone lines with real-time detecting and tracking of intruders, vehicles, or threats [4].

The rapidly increasing demand in this area challenges both academic researchers and industrial practitioners to timely provide analytics theory and system solutions to meet the overwhelming global need. Generally speaking, the challenge is twofold: though hardware of video systems have been fast-developing in the past years thanks to the introduced digital signal processors, hardware-oriented issues are still demanding and unsolved especially for specific applications, system scalability, capability, and real-time performance [5]; on the other hand, algorithm-based analytics have been targeted as the breakthroughs for intelligent video systems and analytics. The state of the art in computer vision and computational intelligence has confirmed that algorithms and software will make a substantial contribution to practical solutions to video analysis and applications in the near future. Hence, this survey is timely to bring the ideas, solutions of the worldwide research community in a summary, to present the latest advances and developments in video systems design, tracking, modeling, behavior understanding, abnormal detection, real-time performance and practical implementation of intelligent video systems and analytics.

There are over 6000 research papers published since 1971, falling into the topics of intelligent video systems and analytics. The concept appeared in early 1970s but the subject emerged to the community around 1980s and developed slowly in that decade, it was well investigated in 1990s and the topic developed further rapidly since 2000. In short, the published papers addressed three aspects of video systems: hardware, software and their applications.

In this paper, we highlight advances in intelligent video systems and analytics in terms of hardware, software and their applications. This survey concentrates the contributions in most recent five years, although some fundamental work has been attracted to researchers as early as in 1970s [6]. For early contributions on video surveillance, please refer to [7].

II. Video Systems

A. Video System Architecture

Intelligent video systems and services incorporate hardware system integration, management, and video processing to endpoint users. A complete solution needs to provide the design,

Manuscript received April 27, 2012; revised September 20, 2012 and January 13, 2013; accepted March 15, 2013. Date of publication April 01, 2013; date of current version August 16, 2013. This work was supported in part by the National Natural Science Foundation of China under Grant 61173096 and Grant ZJNSF R1110679 and the Doctoral Fund of Ministry of Education of China under Grant 20113317110001. Paper no. TII-12-0308.

H. Liu is with the School of Creative Technologies, University of Portsmouth, Portsmouth, PO1 2DJ, U.K. (e-mail: honghai.liu@port.ac.uk).

S. Chen is with the College of Computer Science and Technology, Zhejiang University of Technology, 310023 Hangzhou, China (e-mail: sy@ieee.org).

N. Kubota is with the Graduate School of System Design, Tokyo Metropolitan University, Tokyo 191-0065, Japan (e-mail: kubota@tmu.ac.jp).

Digital Object Identifier 10.1109/TII.2013.2255616

[1][Online]. Available: http://www.icarevision.com

[2][Online]. Available: http://www.infosectoday.com

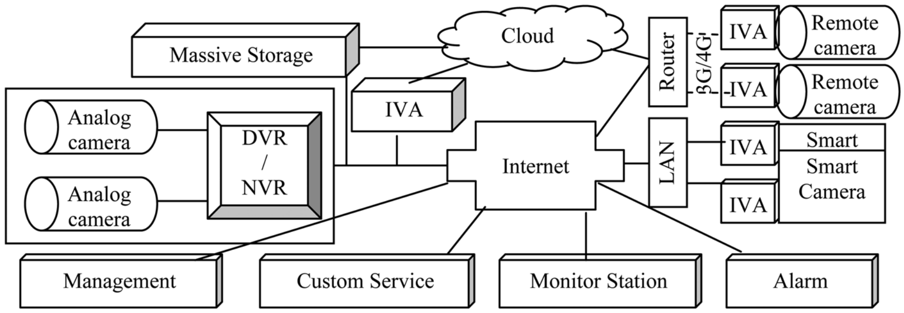


Fig. 1. Overview of the IVS architecture.

integration, installation, and training. System architecture and performance are the main concerns of both providers and users [8]. There are several generations of video systems, i.e., traditional analog systems, analog DVR systems, networked analog DVR systems, networked video encoder systems, networked IP-camera systems, and cloud-based video systems. Fig. 1 illustrates the overview of present video systems. Architectures such as portable video systems with smart phones are also research issues in the community [9], but that is beyond the scope of this paper.

IVS requires analytic processing in either embedded cameras or central servers. In contrast to traditional CCTV camera, high-performance embedded cameras combine video sensing, video processing, and communication within a single device [10]. On the other hand, embedded smart cameras have limited processing power, memory, and bandwidth [11].

To decrease the computational complexity of computer vision algorithms, one way is to achieve low-level image processing at the level of sensor acquisition. Sato *et al.* introduced the design and implementation of a hybrid IVS that consists of an embedded system and a personal computer-based system. The embedded system performs some of the image processing tasks and sends the processed data to the computer [12].

B. Distributed IVS

There are two categories of IVS architectures, i.e., centralized mode and distributed mode [13]. In centralized architectures, video and other information are collected by cameras and brought back to a centralized server for further analysis. In distributed ones, the network cameras are intelligent and are capable of locally processing the video to extract relevant information. On concern of connectivity for city surveillance [14], Internet, cellular networks, and cloud-based video systems are commonly considered to connect the cameras to feed back to the city's central network.

Remote surveillance systems and mobile surveillance systems can be deployed for working in harsh environments. They can be integrated with interoperability solution, providing extended monitoring capability. The systems can also be powered by power cable, self-contained battery power, solar or generator power sourcing. A variety of sensors have been developed to monitor the Earth, ranging from in situ seismographic networks to hyperspectral imaging instruments. Despite an impressive collection of sensing assets, there is still much untapped potential with multiple instruments and remote sensing imagery [15].

Integration of information from multiple cameras is essential in movie production or intelligent surveillance systems [16], [17]. Since complexity can be increased by increasing the number of camera views [11], cooperative [18] and networked [19] strategies are often developed in such IVSs.

In this aspect, Cheng *et al.* present an approach to recognizing driver activities using a multiperspective multimodal (i.e., thermal infrared and color) system for real-time tracking of important body parts [20]. The surveillance task is for robust tracking and profiling of human activities. Huang and Trivedi introduce video arrays for real-time tracking of person, head, and face in a controlled room [21].

There are a number of studies to enable smart video surveillance in a multicamera network [22]. For example, Trivedi *et al.* worked on dynamic context capture and distributed video arrays for intelligent spaces [23]. A collaborative inference framework is presented in [24] for visual sensor networks. An efficient occupancy-reasoning algorithm is used in smart video surveillance based on this kind of framework. The existence probabilities are estimated for every camera, and they are combined using a work-tree architecture with a distributed and collaborative framework.

C. Video Quality Diagnosis

Video systems in poor environments are often associated with poor video quality. Self-awareness of video quality provides a means of diagnosis and alarm for system maintenance. An automatically selected focusing region [25] and digital image stabilization technique [26] are also useful for improvement of video quality. Denoising techniques such as wavelet transform [27] may also be adopted in IVA. Murino *et al.* present an adaptive strategy for regulating the intrinsic parameters of a camera. The parameter regulation procedure is based on the computation of the quality of an acquired signal by means of a set of functions which estimate the image goodness [28].

D. System Adaptability

1) Configuration and Calibration: System configuration and design for wide-area surveillance include modality selection, sensor planning, data fusion, and communication among multiple sensors. An IVS is often required to have far-field view, wide coverage, high resolution, cooperative sensors, and adaptive sensing modalities, sometime with dynamic objects in uncontrolled environments [29]. A smart sensor-based system in [30] integrates a set of analog and digital computing units. The architecture paves the way for a more compact vision system and increases the performance reducing the data flow exchanges with a microprocessor in control [3].

In an IVS, intensive image processing usually demands relevant hardware implementations for real-time performance. However, many applications are hard to be characterized *a priori*, since they take different paths according to events happening in the scene at runtime. Hence, reconfigurable hardware devices are viable, with both real-time performance and dynamic adaptability for the system [31].

System calibration is more related with 3-D computer vision but is sometime also necessary in IVS. A camera calibration using the geometry properties of road lane markings is proposed in [32]. The camera parameters include pan angle, tilt angle, swing angle, focal length, and camera distance. Results show that the method outperforms the others in terms of accuracy and noise sensitivity. For adaptive camera regulation for investigation of real scenes, Murino *et al.* analyze the feature histograms to compute actual camera parameters, i.e., focusing distance, aperture diameter, electronic gain, and black level [28].

2) Capability and Scalability: Cross-platform capability and scalability are important features pursued in the IVS community. Many studies treat central processing approaches with scene analysis processed inside a central server. Such approaches require tremendous efforts in building the system and limit the scalability. To accomplish scalable IVSs, an inference framework in visual sensor networks is necessary, one in which autonomous scene analysis is performed via distributed and collaborative processing among camera nodes without necessity for a high-performance server [24]. IVSs also require reliable transmission of high-quality video over networks using available resources. Scalable video coding is a video compression technology to support potential capabilities [33]. Ghiasi *et al.* exploit reconfigurable hardware devices embedded in a number of networked cameras [31].

E. Data Management and Transmission

IVSs collect, index, store, and deliver video 24/7 and allow users to monitor scenic events in real time. Along with the rapidly increasing mass of online and offline videos, it is demanding to develop efficient methods for management and retrieval of video clips/segments based on the semantic content [34]. Hong *et al.* describe an intelligent video categorization engine that uses the learning capability of artificial neural networks to classify suitably preprocessed video segments into a predefined number of semantically meaningful events [35]. Tsai *et al.* design a videotext in a picture display system which can extract the videotexts in the subchannel and then combine these videotexts with the main channel [36]. Single-instruction multiple-data-based mechanisms were created to enhance the computational efficiency on numerous convolutions and accumulations in videotext extraction.

Managing imprecise queries using semantic content is dramatically harder than queries based on low-level features due to the absence of a proper continuous distance function, which is an active open research area [37]. Kim and Shibata present a natural language approach to content-based video indexing and retrieval [38]. The experiments illustrate the proposed index structure has superior retrieval capabilities compare with those used in conventional methods. The task also relates video structuring, object, and text detection [39], [40], as well as visual saliency analysis [41].

For reliable transmission of high-quality video in situations of web-based monitoring, as the live data through the packet-switched network environment of the Internet can result in packet loss and quality degradation. Grgurevic *et al.* attempted the possibilities of insuring the credibility and authenticity of the surveillance video by digital signing, using the public key infrastructure as part of inter-operable traffic and information system in the future intelligent transport-systems [42].

A summary of some early works on intelligent surveillance in visible and multimodal framework can also be found in [29] and [43].

III. Analytic Tasks

A. Object Analytic Attributes

1) Target Description: The analytic methods are heavily overlapped with those in computer vision and image processing. The workflow and the most important steps in video surveillance usually include background subtraction, moving foreground segmentation, object/shadow detection, tracking and classification, and event recognition [44], [45]. In every step of IVA, the target in the video sequence has to be represented by some mathematically describable features. Common descriptors developed in the community include local color histogram, texture, and geometrical shapes. On this issue, researchers have been searching methods such as invariant moments, parametrical geometry, wavelet, multiresolution representation, statistical and structural approaches, edge histogram, tangent space, hidden Markov model (HMM), and centroid-radii model.

The well-known RANSAC method is applied to estimate parameters of a mathematical model of geometrical shapes, parametrical geometry, and others. In particular, various types of modified SAC algorithms have been proposed to improve the tradeoff between computational cost and accuracy in parametrical geometry. For example, MSAC is more robust than RANSAC, because MSAC uses not only the number of outliners but also the error of candidate pairs included in Inliers. On the other hand, the method using genetic algorithm sampling consensus (GASAC) can improve the computational efficiency and accuracy in IVA.

2) Target Detection: Target detection and recognition by contextual information contributes a part of analytic tasks in IVS. It is well investigated to detect and classify traffic signs from road images in real time as a support tool for guidance and navigation of intelligent vehicles [46], but the hundreds of traffic sign types and their various shapes and colors make it still difficult to develop a generalized method of traffic sign detection [47].

Background subtraction is crucial for detecting moving objects in videos. For instance, to identify the entrance legality in a restricted area, a background subtraction technique is used in [48], to detect moving objects. Obtaining foreground regions by subtracting its background image can be achieved by different methods, such as the mixture of Gaussians [49], the kernel density estimation, or the concurrence of image variations.

Instead of using a frame-subtraction method, Gao *et al.* introduce a Marr wavelet, kernel-based background modeling method, and a background subtraction method based on binary discrete wavelet transforms [50]. Shadow removing is a difficult task in a complex context. Li *et al.* attempted to use a Gaussian mixture model (GMM) for background removal and detection of moving shadows [51]. Two indices are defined for characterizing nonshadowed regions where one indicates the characteristics of lines; the other index can be characterized by the information in gray scales of images which helps us to build a newly defined set of darkening ratios based on Gaussian models.

3) Target Tracking: IVA often needs detection and tracking of suspicious objects [11]. Tsai *et al.* propose a dissimilarity measure based on the optical-flow technique for surface defect detection aiming [52]. Chen *et al.* develop a hierarchical background model based on segmented background images with convincing experimental results [53].

Object tracking aims to detect the path of objects moving by obtaining input from a series of images [54]. It tracks the object by assuming the initial state and noise covariance. Wang *et al.* present an efficient approach to embed hyperspectral imaging capability in an intelligent panoramic scanning system for real-time target tracking and signature acquisition [55].

For intelligent traffic monitoring, lane estimation and moving object tracking are key technologies to success [50], [56]. Liu and Yung construct a tracking cue as a weighted log likelihood ratio by both the object GMM and its surrounding GMM [57]. It works even if the object appearance varies due to changes in viewing angle, scale, and illumination. Kwak *et al.* proposed detection-abandoned luggage with an intelligent surveillance system for public places [44]. To recognize an abandoned luggage event, a finite state automaton is constructed in which each state represents a certain luggage status.

A monitoring method based on the cellular model is proposed by Hsu *et al.* [58] to monitor human activities in the indoor environment. Guan *et al.* argue that automatic detection of human faces needs to combine feature extraction and face detection based on local normalization, Gabor wavelets transform and Adaboost algorithm [59]. An efficient sequential approach is proposed in [4] and [60] to track multiple objects through crowds in real-time IVSs. Morioka *et al.* worked on a cooperative method for adaptive camera selection for target tracking in multicamera system [18].

4) Real-Time Analytics: IVA algorithms often work in a real-world environment and at real-time speed [44], in spite of offline operation in a central analytics server. IVA includes a wide variety of functions, e.g., idle object detection, trajectory tracking, and spatial video denoising [27] in live image sequences. Events should be immediately displayed for triggering corresponding alerts, while the video clip can be stored for later review. Zhang *et al.* employs fuzzy genetic algorithm to boost the computing efficiency of covariance matching for optimal solution in a large image region [61]. A real-time system is also used in [62] for detecting tailgating, an example of complex interactions and activities within a vehicle parking scenario, using an adaptive background learning algorithm and intelligence to overcome the problems of object masking, separation, and occlusion.

Furthermore, we must reduce computational time as much as possible in image processing. General-purpose graphics processing unit (GPGPU) has been applied feature extraction, e.g., scale-invariant feature transform (SIFT) and object tracking in real-time. On the other hand, the standard SURF is several times faster than SIFT.

B. Motion Pattern Recognition

Motion detection is extremely important in IVS. Automatic interpretation of human and vehicle motion in surveillance videos is inevitable to detect abnormal behaviors [63]. Based on the motion property of the dynamic background and that of the moving vehicles, Zhang *et al.* present an adaptive motion histogram for vehicle segmentation. The algorithm consists of two procedures: adaptive background update and motion histogram-based vehicle segmentation [64]. Vanne *et al.* propose a configurable motion estimation architecture for a variety of fast block-matching algorithms. The experiments reveal that the proposed implementation is able to process real-time fixed block-size motion estimation at full HDTV resolution.

Zhou *et al.* develop an automated activity analysis and summarization for eldercare. They construct a silhouette extraction, human detection, and tracking algorithm for indoor environments. Important activities of daily living statistics are extracted for automated functional assessment [65]. A frequent trajectory patterns mining algorithm is proposed in [66] to learn the object activities and classify the trajectories in IVSs. The distribution patterns of the trajectories were generated by an *a priori*-based frequent patterns mining algorithm and the trajectories were classified by the frequent trajectory patterns generated. An approach is proposed in [67] for learning sequenced spatiotemporal activities in outdoor traffic intersections.

C. Behavior and Event Analysis

In public venues, we cannot stop terror or crime by using a virtual perimeter in the video. Site surveillance cannot rely solely on motion to identify threats, but video content analysis must be carried out for abnormal behavior detection or event recognition. Behavior and event recognition detect and track objects in video images, applying a set of rules to detect crimes or violations such as abandoned objects, stopped cars, object removal, excessive speed, abnormal moving, crowding, loitering, or stalking. In particular, for complexity analysis of human behavior in a moving crowd, extraction of object velocity and position can help mathematical modeling of crowd dynamics [68].

Atallah and Yang summarized three different types of behavior patterns, i.e., a sequence of clearly defined consecutive activities, concurrent activities occurring at the same time, and subactivities belonging to three activity classes occurring concurrently and interleavingly [69]. Behavior modeling and activity recognition can then be carried out by analyzing these patterns using methods like HMM [14], [70].

In this area, Albusac *et al.* [71] propose a conceptual framework based on normality analysis to detect abnormal behaviors by means of normality concepts. Normality specifies how a certain object should ideally behave in an environment. The definition of the normal path concept is studied in order to analyze behaviors in an outdoor environment. In [66], a fuzzy c-means-based learning algorithm and a mean-shift-based clustering procedure were used to construct the representation of trajectories. The algorithm can be used to describe activities and identify anomalies.

Behavior and event recognition are widely used in many surveillance tasks, e.g., intrusion detection (entering a restricted area) [72]. For neurological diagnosis and management, analysis is applied to classification of movement disorders [73]. In development of an intelligent emergency response system in [74], some volunteers are asked to assume a series of postures, e.g., walking/standing, sitting/lying down, stooping, stretched lying, and tucked lying. These tasks are simulated to detect falls indoor [75].

IV. Analytic Methods

This paper mainly concerns the analytic methods that are specially applied in IVS. Selected analytic methods are summarized as follows.

A. Intelligence

IVA have attempted to apply all adaptive and intelligent methods of neural network [76], genetic algorithm, knowledge-based approaches [71], particle filtering [50], particle swarm optimization [4], finite state automation [44], reasoning [24], self-organizing maps (SOMs) [46], support vector regression (SVR) [77], Kalman filtering [78], semantic analysis [71], Markov models [79], decision tree [65], and clustering [66]. Specially, an online neural estimator is proposed in [80] for object tracking and fixation. Hong *et al.* compare the performance of two neural networks: Kohonen's SOM and fuzzy adaptive resonance theory (Fuzzy ART) [35].

For adaptive detection, a normalization technique is incorporated in [59] by local histograms with optimal adaptive correlation technique, so that it can avoid inconsistent performance caused by sensitivity of variation illuminations such as local shadowing, noise, and occlusion. The approach uses a cascade of classifiers to adopt a coarse-to-fine strategy for achieving higher detection rates with lower false positives. In [58], the indoor area is divided into several unit areas in which each unit is considered as a simple cell in the cellular model. A rectangular box is used to group those neighboring active cells into a unit to represent a moving object. They further apply the gray relational analysis to detect and track multiple moving objects. Bio-inspired adaptive hyperspectral imaging for real-time target tracking is also attempted in [55].

B. Cooperative and View Selection

In multicamera systems, there is a problem of selecting the right view to display among the multiple video streams. A view is defined by the camera index and the parameters of the image cropped within the selected camera. View selection is often required when selective attention [23], [94] or surveillance at strategic locations are implemented [17].

A cooperative multicamera system allows a single operator to monitor activities in a cluttered environment using a distributed network of video sensors [11], [16]. Video understanding algorithms are developed to automatically detect people or vehicles and seamlessly track them using a network of cooperating active sensors [19], [54]. Wang *et al.* present a wireless embedded smart-camera system for cooperative tracking and detection of composite events spanning multiple camera views [11]. Instead of transferring or saving every frame or trajectory, events of interest are detected. Simpler events are combined in a time sequence to define semantically higher level events.

A collaborative and dynamically adaptive tracking system is introduced in [31]. It is justified that dynamic adaptation of the system is necessary through scenarios and applications.

In [18], a fuzzy automaton-based camera-selection method is introduced. The camera-selection decision is driven by fuzzy automaton based on the previously selected camera and the tracking level of the object in each available camera. The results show that the method is efficient for adaptive camera selection in multicamera environment and helps easy construction of multicamera placement.

Chen and de Vleeschouwer propose criteria for optimal planning of viewpoint coverage and camera selection. Perceptual comfort is discussed as well as efficient integration of contextual information, which is implemented by smoothing generated camera sequences to alleviate flickering visual artifacts and discontinuous storytelling artifacts [16].

A fully digital autofocusing system is presented in [25] with automatic focusing region selection and *a priori* estimated dataset of circularly symmetric point-spread functions. The approach provides realistic, unsupervised blur estimation by analyzing the entropy and edge information in the automatically selected focusing region.

C. Integration and Statistics

Statistics may be applied for event detection, counting, routing, guidance, surveillance, and flow control. Here, the flow mainly includes crowd flow and traffic flow. Besides the traffic flow detection [81] that is going to discussed in detail later, there have been many works on attention control and statistics for business, tourism, public safety, civilization, exhibition, and markets/shops [14].

Integration of information from multiple sources or cameras is necessary to make the system more intelligent [29]. Dynamic synthesis of virtual views is developed for observing the environment from arbitrary vantage points. Takacs describes a real-time image processing and sensor fusion system for aerial vehicles in need of autonomous landing, guidance, and obstacle avoidance. The system can process and display information merged from multiple image sources including a high-resolution millimeter-wave radar, a stored terrain with 3-D airport database, FLIR, as well as visual reference images [82].

D. Networked Analytics

Many applications perceive visual information through networks of embedded sensors [31]. Distributed smart cameras perform real-time computer vision thanks to a confluence of simultaneous advances in disciplines of computer vision, image sensors, embedded computing, and sensor networks. For instance, in [21], a network of video cameras is used for person tracking in an intelligent room. Comaniciu *et al.* present a real-time foveation system for remote and distributed surveillance. The system performs detection, tracking, selective encoding, and efficient data transmission. A client-server architecture connects a radial network of camera servers to their central processing unit [22].

E. Learning and Classification

Learning and classification are powerful for object detection and event recognition [62]. An adaptive learning method is used in [65] to estimate the physical location and moving speed of a person. Hierarchical decision tree and dimension reduction methods are explored for human action recognition.

To detect moving objects in image sequences and identify the entrance legality in a restricted area, Shih *et al.* extract three object features, i.e., the position, the size, and the color, to track the detected entrants [48]. After that, the entrant was segmented into three parts for locating the region of interest (ROI) using a watershed transform. Dominant color features extracted from the ROI are classified for preventing the illegal entrance. Kafai presented a stochastic multiclass vehicle classification system that classifies a vehicle into one of four classes: sedan, pickup truck, SUV/minivan, and unknown [83].

Detecting anomalies exhibited in complex behaviors which are subtle and difficult to owing to the complex temporal characteristics and correlation among multiple objects' behaviors [84]. In [85], anomaly detection is achieved by the combination of the normalized log-likelihood with respect to the first-stage HMM and that to the second-stage multi-observation HMM, which are determined from the computation of marginal probabilities in a filtering process. A threshold is used for judging. The results outperform existing methods in detecting durational anomalies.

Computer vision system can be inspired by the human visual system for organizing the different visual routines that need to be carried out [86]. Wang *et al.* use bio-inspired adaptive hyperspectral imaging for real-time target tracking [55]. Quek *et al.* propose a brain-inspired neural cognitive approach to SARS thermal image analysis [87]. From the study by Callan *et al.* [88], neural substrates of cost-weighted decision making can be assessed by investigation of driver's decision making. They investigate neural correlates of resolving uncertainty in driver's decision making.

F. 3-D Sensing

3-D reconstruction from 2-D images has been the important foundation for solving the problems in robotics and computer vision fields [89], [90]. Among important issues in IVS, one of the main problems is to localize the object features in the 3-D space. Technologies such as stereo vision [56], [76], [80], laser scanning, or structure from video motion are common sensing methods to obtain 3-D information.

Habib develops an intelligent fiber-grating-based 3-D vision sensory system that enables real-time object detection, monitoring, and tracking [91]. Gavrila and Munder propose a method with tight integration of several consecutive modules, i.e., (sparse) stereo-based ROI generation, shape-based detection, texture-based classification, and (dense) stereo-based verification [92].

In recent years, various types of real-time 3-D modeling methods using RGB-D cameras have been adopted after Microsoft developed the Kinect sensor [54]. These methods estimate parameters of a mathematical model by choosing and matching correct pairs of corresponding points between two images at t and $t - 1$ based on features such as SIFT and SURF. A 3-D model is then built using distance information by coordinate transform based on the estimated parameters.

V. Applications

Fundamentals such as system design, data management, video processing, calibration, edge enhancement, background subtraction, recognition/detection, tracking, and motion understanding are often the common techniques useful for all IVS application. Although some works are application-specific in the presentation, they may also be used for general purposes because of the generality of the method, e.g., the cellular model proposed in [58] to monitor human activities and normality analysis to detect abnormal behaviors [71].

A. Management

IVS provides an efficient means to many management tasks, by applying the technologies of automatic people counting, access control, flow control, and attention control [14]. Requirements are arisen from managers of all fields for day-to-day information processing and decision making [3]. The business needs or outside regulations currently are mainly from campus, government, retail, airport, seaport, commercial office, gaming, banking, gathering event, industrial, residential, etc. For example in a store or supermarket, IVS is for not only securing inventory from theft and ensuring that every transaction is complete and legitimate, but also improving the service by statistical analysis of customers' shopping manner and optimizing employee productivity in retail.

B. Traffic Control and Transportation

IVA have been widely applied in traffic control and transportation, e.g. lane traffic counts, illegal U-turn, illegal lane change, wrong direction, wrong way, vehicle requiring assistance, incident detection, etc.

IVS is important for gathering data for *intelligent transportation system* applications over a traffic flow by detection of moving vehicles [64], [93]. An automatic particle filtering algorithm is used in [50] to track the vehicle and monitor its illegal lane changes. Li *et al.* propose an efficient algorithm for removing shadows of moving vehicles caused by nonuniform distributions of light reflections in the daytime [51]. Traffic-flow-detection technology includes the use of a loop detector, an infrared detector, an image detector, and a microwave detector. Wang *et al.* propose a channel awareness vehicle detector that employs only one pair of transmitter–receiver antennas to simultaneously perform the multilane and multi-vehicle identifications. By using the characteristics of channel variations, the vehicle detector can determine the vehicle location, speed, and type [81].

With a machine learning framework, Chen *et al.* attempt real-time traffic density estimation by using a hidden Markov models to probabilistically determine the traffic density state [79]. Automatic incident detection and protection are attracting in many cities for detecting traffic incidents to provide smoother, safer and congestion free traffic flow. Cai *et al.* propose a robust real-time algorithm to detect snow movement in video streams [94].

In heavy traffic flow conditions, vehicles have limited maneuverability which affects the magnitude of response to incident-induced traffic disturbances and how fast changes in these traffic variables can signal the occurrence of an incident. Such characteristics are usually used to formulate a loop-based algorithm. Mak and Fan reported two video-based automatic incident detection algorithms, the individual detection evaluation and combined detection evaluation algorithms [95]. The algorithms are developed for the detection of lane-blocking incidents in heavy traffic flow conditions using the Central Expressway in Singapore. The algorithms detect incident-induced traffic speed and occupancy disturbances differently: the former processes information at each individual detector station and the later processes information at two adjacent detector stations.

Detecting license plates is crucial and inevitable in the vehicle license plate recognition system [96]. It is required to ensure consistent and reliable plate capture for practical applications. There are several situations of plate detection, i.e. general license plate capture at daytime, close range license plate capture, license plate capture at night or in the dark environment, and overview camera to capture both plates and vehicle details. It is found to improve the recognition performance in terms of speed by rapidly scanning input images focusing only on ROIs, while at the same time it does not reduce the system effectiveness.

C. Intelligent Vehicle

1) Pedestrian detection: IVA can be applied to intelligent vehicles for traffic sign recognition, people tracking, and driver assistance. Pedestrians are the most vulnerable participants in traffic. The first step toward protecting pedestrians is to reliably detect them in a real-time framework [97], [98]. An actual road test shows that the algorithm can effectively remove the influence of pedestrians and cyclists in the complex environment and can track the moving vehicle exactly. An approach is presented in [77] for pedestrian detection in urban traffic conditions using a multilayer laser sensor mounted onboard a vehicle.

2) Driver-Assistance System: Driver-assistance systems that monitor driver intent, warn drivers of lane departures, or assist in vehicle guidance, collision avoidance, and autonomous driving are all being actively studied in the community.

In the study of neural correlates of resolving uncertainty in driver's decision making, turning right in left-hand traffic at a signalized intersection is simulated in [88] by graphic animation-based videos. When the driver's view is occluded by a big truck, the uncertainty of the oncoming traffic is resolved by an in-car video assist system that presents the driver's occluded view. An attempt for driver body tracking and activity analysis is proposed in [20].

Surround information or maps can help in studies of driver behavior as well as provide critical input in the development of effective driver assistance systems. Gandhi and Trivedi focus on the capture of vehicle surroundings using video inputs [56].

3) Traffic Sign Detection and Recognition: Roadway signs are important for safety, and transportation agencies need to identify sign condition changes to perform timely maintenance. Tsai *et al.* propose an algorithm to detect three condition changes: missing, tilted, and blocked signs, using GPS data, and video log images [99].

An innovative image processing model is proposed in [47] to automatically detect traffic signs and dramatically reduces the sign inventory workload. The method is composed of: 1) a generalized traffic sign model to represent the entire class of traffic signs; 2) a statistical traffic sign color model; 3) a traffic sign region of interest detection system using polygon approximation; and 4) traffic sign candidate decision rules based on shape and color distributions. Prieto and Allen alternatively propose a method for the detection and recognition of traffic signs using self-organizing maps [46]. The method first detects potential road signs by analyzing the distribution of red pixels within the image and then identifies these road signs from the distribution of dark pixels in their pictograms. A fast and robust framework for incrementally detecting text on road signs is presented in [100].

In other aspects, there are many contributions on IVS for transportation, such as annotating traffic videos [101] and authenticity and credibility of videos [42]. Surface movement guidance and control is introduced in [102], which provides routing, guidance, surveillance, and control to aircraft and vehicles [82]. An intelligent video surveillance system for aircraft is introduced in [17], which allows the surveillance data to be stored within the aircraft and monitored by one of the flight crew. The monitoring crew will be responsible for identifying the anomaly within the aircraft and take necessary preventive actions.

D. Healthcare and Life Sciences

Video image analysis is able to provide quantitative data on postural and movement abnormalities and thus has an important application in neurological diagnosis and management [73]. Quek *et al.* investigate the application of several novel brain-inspired soft computing techniques in the study of the correlation of superficial thermal images against the true internal body temperature [87].

Automated IVS surveillance is proposed to ensure safety of the elderly while respecting privacy and the topic becomes interesting but with many challenges. Fleck and Strasser reported a prototype 24/7 system installed in a home for assisted living for several months and shows quite promising performance [10]. Zhou *et al.* study how IVS and IVA can be used in eldercare to assist the independent living of elders and to improve the efficiency of eldercare practice [65].

To help blind people for navigation without collision, Nagarajan *et al.* propose a real-time scheme in providing vision substitution to visually handicapped people [103]. The system has a computer, a headgear with a digital video camera, and a set of stereo earphones interconnected. The software processes the image to locate the object and its boundaries and the result is mapped onto structured stereo sound patterns, so that the blind can understand the environment around him through the set of stereo earphone.

Sensors and intelligent specialist software are helping biologists by improving the selectivity of images captured and stored, and the responsiveness of remote systems to their live imaging needs. Automated and tele-operated equipment greatly increases observation potential whilst avoiding the disturbance of human presence [104].

In the future, IVS is expected further for wide applications in studying life sciences which involve scientific investigation of living organisms, such as plants, animals, and human beings, as well as related research for ethology, ecology, zoology, wildlife biology, systems biology, health sciences, biodynamics, environmental science, evolutionary biology, and biomedical sciences [105].

E. Security and Military

Security and military applications have been the primary motivation in developing IVS [29]. From small outposts to military cities with thousands of people, everything is a potential target. Security systems are ready to adapt as needed. IVS can offer customized, mission-critical security applications and systems that can adjust quickly to changing needs and government regulations. A basic security application of IVS is for people detection and tracking in an environment [19], [54]. Based on this, the system may be employed to estimate the number of accesses in public buildings as well as the preferred followed routes [106]. Such an automatic surveillance system is developed in [45] to detect several dangerous situations in subway stations.

Intrusion detection is often required for perimeter protection. It may take several forms, such as crossing a boundary to move into the site, moving in a no-man zone, or throwing an object over a fence. In [48], the color features of an employee's uniform were extracted to identify the entrance legality in a restricted area of an open space. It is reported that the one Chilean salmon farm experiences up to $2 million a year in theft of salmon and salmon eggs. An IVS is set up for detecting intrusion by boats and unauthorized persons, while differentiating between marine predators and harmless small animals entering the perimeter [72].

Leed *et al.* apply the IVS technology, focusing on a specific proposal to combat the modern scourge of missile threat to civil aviation [107]. IVSs integrated with high-value video surveillance equipment and computer-aided dispatch servers are proposed as a methodology to detect and dispatch effective preemptive responses to the threats of shoulder-fired missiles directed against commercial airlines operating out of airports in densely populated areas. Abandoned object detection is necessary for security because unattended or abandoned luggage can be used as a means of terrorist attack, especially for bombs [44].

Control of sensitive access is required for monitoring access points such as doors to protect unauthorized entrance and exit. IVA is used to prevent or detect some tactics of abnormal access. A scalable access point monitoring solution can be constructed with embedded video detection and transmission devices for tailgating detection, combined with a central IVA management system. Detection accuracy can be enhanced by using multiview analysis, with additional viewing angles for alarm verification and optionally alert upon certain behaviors near the access point.

VI. Research Directions and Discussions

While IVS and IVA have been developed as useful approaches for many applications, some problems still exist in its adaptation in practical environments. The intelligent surveillance systems are still not widely deployed in practical applications. Reliability and accuracy may be the main reason of limitations in the current systems. Not only simply applying for detection and tracking, researchers are making efforts to improve the methods mainly in the following aspects.

A. Real-Time Application and Computational Complexity

IVS often requires real-time processing for quick responses. Analytics have to be performed at the frame rate of the video system, e.g., 30 fps [108]. However, computer vision often takes very complex computations. The computational complexity restrains the systems of real-time application. Nevertheless, IVA can take all possible solutions, sometime at any cost, for practical implementation of security systems, while computer vision mainly focuses on theoretical advances for low cost and high accuracy. Therefore, it can often be advised to use advanced computers, GPU, and parallel computing.

B. Reliability and Flexibility

Reliability and flexibility are mostly concerned in practical applications. A good IVS should be flexible in different environments and weather conditions, with varying illuminations, initial parameters, predefined conditions, as well as rain, snow, shadows, noises, and occlusions. For example, spatial-temporal distribution can be analyzed for robust foreground segmentation and feature extraction. It would be interesting in the community for developing a method to automatically determine the number of frames for any application so that both spatial and temporal information can be combined for optimization.

C. Efficiency and Accuracy

Both efficiency and accuracy are important factors in IVA. Unfortunately, they cannot always be satisfied with both sides. In all analytic methods of object detection, tracking, background subtraction, 3-D sensing, learning, classification, video text detection, behavior detection, and event recognition, there are still many challenges to achieve better accuracy [39], [40]. IVS is often required with extensive computation of visual information, but also given the detection task of objects or events. Even without considering the efficiency, the accuracy of current IVSs is still very low. Those systems often suffer from high false alarm rates due to environmental uncertainties when they operate in an outdoor environment.

D. Distributed and Networked

Nowadays, there are hundreds of thousands of cameras distributed any one of many cites. Distributed smart cameras represent important future IVS systems. However, networked and centralized analysis and control become more important than ever for efficient use of all of the video sources. Simultaneously, factors of coding, encryption, packetization, authentication, and transcoding have to take into consideration for data security in the networks. One attempt of credibility and authenticity of digitally signed videos in traffic is found in [42]. It is realized by comparison of the hash values of the frames stored in the database of the surveillance centre with the values obtained from the interested subjects.

E. Standards and Performance Evaluation

Standards in IVA and IVS become obviously important now due the rapid increasing of related products and markets. There are rare standards available yet in the field. For applications as successful as by traditional surveillance, it is crucial that IVS should be deployed easily, without requiring computer vision expertise to customize them for every installation. Since late 2008, the Open Network Video Interface Forum (ONVIF) becomes a global open standard for the interface of physical IP-based security products, which can achieve interoperability between network video products regardless of manufacturer. Another competitive standard released almost at the same time is Physical Security Interoperability Alliance (PSIA), which make efforts for security system and device integration to be as simple as the "plug and play" interoperability. Venetianer and Deng recently discussed some of the major challenges involved in software testing [3]. They provide a concept of utilizing fuzzy evaluation to handle boundary conditions. The standards are necessary not only for hardware, but also for software performance. The Imagery Library for Intelligent Detection Systems (iLids) is the U.K. government's standard for video detection [8]. The first four iLids scenarios were released in November 2006, and there have been annual evaluations for these scenarios since then.

F. Hardware and Software for Video Processing

Researchers are pursuing new generation of hardware devices and software strategies so that artificial vision can understand much of what biological vision can. Kohler *et al.* recently designed a smart camera with an array of elementary motion detectors, where the motion-detection directions and the angle between correlated receptors are reconfigurable online [109]. It allows a flexible and simultaneous detection of complex motion fields such as translation, rotation, and zooming. The compact device benefits many motion-based applications such as obstacle avoidance, distance control, and speed regulation.

G. Information Fusion and Cloud Computing

Alignment and fusion of different media sources are often required for an IVS, e.g., audiovisual information fusion [110], human–computer interfaces, and multiple cameras and instruments. Krotosky and Trivedi present an analysis of color-, infrared-, and multimodal-stereo approaches to pedestrian detection [98]. Traditional video systems require infrastructures including expensive servers to store and process a huge amount of videos. Using cloud computing to collect and process multimedia streams can benefit on many points, e.g., cross terminal, cross content, adaptive to different networks, integration of intelligent processing algorithms, resource integration, sharing, and easy to video searching and is especially useful for real-time IVS for large scenes or intelligent city domain [111], [112].

H. Cyberphysical System (CPS)

IVS and IVA provide effective means of sensing in intelligent environments [113], which is much related with the recent concepts of CPS or Internet of Things (IoT). The former describes a system featuring a tight combination and coordination between computational and physical elements, while the latter refers to the networked interconnection of everyday objects whose purpose would be to make all things communicable. Here, IVA can be a part of ambient intelligence for either IoT or CPS [105], [114]. Especially in embedded systems, the emphasis tends to be more on the computational elements (IVA) and less on physical elements (IVS). A fully embedded video CPS (IVS+IVA) is typically designed as a network of interacting elements with multimedia input and output instead of as standalone devices. The video obtained by IVS might be combined with other information from the IoT.

I. More Intelligence

Integration of IVA with artificial intelligence methods can certainly yield better performance. Actually, fuzzy logic, neural network, and genetic algorithms have been attempted for resolving a complex task in whole. However, the available artificial intelligence itself is still at a relatively low level, which affects much in improving IVA performance. Computer intelligence might be always imperfect for vision understanding, and thus it is important to incorporate human knowledge in IVS. Of course, more intelligence of video processing is ever-increasingly expected for IVA in the future.

VII. Conclusion

This paper has summarized the recent development of intelligent video systems and analytics. Typical contributions are addressed for a variety of applications. Representative works are listed for readers to have a general overview of the state-of-the art. A bundle of methods are investigated in regard to solutions of video analysis, including video system design, real-time analysis, detection, tracking, background subtraction, learning, classification, 3-D sensing, motion and behavior detection, and event recognition. Future research challenges and directions have been outlined in the end. In short, the current video systems might not be as good as people expect at the moment, the applications of the technology however do spread rapidly. Thousands of companies, emerged in the past decade, are especially developing products of intelligent video systems driven by both academic and industrial demands. It is expected that the IVS would be maturely applied rapidly in this decade.

References

[1] H. Liu, S. Chen, and N. Kubota, "Guest editorial special section on intelligent video systems and analytics," *IEEE Trans. Ind. Inf.*, vol. 8, no. 1, p. 90, Feb. 2012.

[2] D. Elliott, "Intelligent video solution: A definition," *Security*, pp. 46–48, 2010.
[3] P. L. Venetianer and H. L. Deng, "Performance evaluation of an intelligent video surveillance system—A case study," *Comput. Vis. Image Understanding*, vol. 114, no. 11, pp. 1292–1302, 2010.
[4] X. Zhang, W. Hu, W. Qu, and S. Maybank, "Multiple object tracking via species-based particle swarm optimization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 11, pp. 1590–1602, Oct. 2010.
[5] G. Wang, L. Tao, H. Di, X. Ye, and Y. Shi, "A scalable distributed architecture for intelligent vision system," *IEEE Trans. Ind. Inf.*, vol. 8, no. 1, pp. 91–99, Feb. 2012.
[6] M. J. Streibel, "Dialog design and instructional-systems design for an intelligent videodisc system," *Videodisc and Opt. Disk*, vol. 4, no. 3, pp. 216–229, 1984.
[7] H. M. Dee and S. A. Velastin, "How close are we to solving the problem of automated visual surveillance? A review of real-world surveillance, scientific progress and evaluative mechanisms," *Mach. Vis. Applications*, vol. 19, no. 5–6, pp. 329–343, 2008.
[8] K. Sage, A. Nilski, and I. Sillett, "Latest developments in the iLids performance standard: New imaging modalities," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 25, no. 7, pp. 4–11, Jul. 2010.
[9] R. Cucchiara and G. Gualdi, "Mobile video surveillance systems: An architectural overview," *Mobile Multimedia Process.: Fundam., Methods, Applications*, vol. 5960, pp. 89–109, 2010.
[10] S. Fleck and W. Strasser, "Smart camera based monitoring system and its application to assisted living," *Proc. IEEE*, vol. 96, no. 10, pp. 1698–1714, Oct. 2008.
[11] Y. Wang, S. Velipasalar, and M. Casares, "Cooperative object tracking and composite event detection with wireless embedded smart cameras," *IEEE Trans. Image Process.*, vol. 19, no. 10, pp. 2614–2633, Oct. 2010.
[12] K. Sato, B. L. Evans, and J. K. Aggarwal, "Designing an embedded video processing camera using a 16-bit microprocessor for a surveillance system," *J. VLSI Signal Process. Syst. Signal Image Video Technol.*, vol. 42, no. 1, pp. 57–68, 2006.
[13] F. Nilsson, "Making a business case for 'edge' network intelligence," *Security*, vol. 58, p. 56, 2009.
[14] K. F. MacDorman, H. Nobuta, S. Koizumi, and H. Ishiguro, "Memory-based attention control for activity recognition at a subway station," *IEEE Multimedia*, vol. 14, no. 2, pp. 38–49, 2007.
[15] H. Li, H. Gu, H. Yanshun, and Y. Jinghui, "Object-oriented classification of high-resolution remote sensing imagery based on an improved colour structure code and a support vector machine," *Int. J. Remote Sens.*, pp. 1453–1470, 2010.
[16] F. Chen and C. De Vleeschouwer, "Personalized production of basketball videos from multi-sensored data under limited display resolution," *Comput. Vis. Image Understanding*, vol. 114, no. 6, pp. 667–680, 2010.
[17] N. Thanthry, I. P. Emmadi, A. Srikumar, K. Namuduri, and R. Pendse, "SVSS: Intelligent video surveillance system for aircraft," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 24, no. 10, pp. 23–29, Oct. 2009.
[18] K. Morioka, S. Kovacs, L. Joo-Ho, and P. Korondi, "A cooperative object tracking system with fuzzy-based adaptive camera selection," *Int. J. Smart Sens. Intell. Syst.*, pp. 338–358, 2010.
[19] R. T. Collins, A. J. Lipton, H. Fujiyoshi, and T. Kanade, "Algorithms for cooperative multisensor surveillance," *Proc. IEEE*, vol. 89, no. 10, pp. 1456–1477, Oct. 2001.
[20] S. Y. Cheng, S. Park, and M. M. Trivedi, "Multi-spectral and multi-perspective video arrays for driver body tracking and activity analysis," *Comput. Vis. Image Understanding*, vol. 106, no. 2–3, pp. 245–257, 2007.
[21] K. S. Huang and M. M. Trivedi, "Video arrays for real-time tracking of person, head, face in an intelligent room," *Mach. Vis. Applic.*, vol. 14, no. 2, pp. 103–111, 2003.
[22] D. Comaniciu, F. Berton, and V. Ramesh, "Adaptive resolution system for distributed surveillance," *Real-Time Imaging*, vol. 8, no. 5, pp. 427–437, 2002.
[23] M. M. Trivedi, K. S. Huang, and I. Mikic, "Dynamic context capture and distributed video arrays for intelligent spaces," *IEEE Trans. Syst. Man Cybern. A, Syst. Humans*, vol. 35, no. 1, pp. 145–163, Jan. 2005.
[24] C. Yongil, O. L. Sang, and S. Y. Hyun, "Collaborative occupancy reasoning in visual sensor network for scalable smart video surveillance," *IEEE Trans. Consum. Electron.*, vol. 56, no. 3, pp. 1997–2003, Aug. 2010.
[25] J. Jaehwan, Y. Inhye, K. Donggyun, L. Jinhee, and P. Joonki, "Fully digital auto-focusing system with automatic focusing region selection and point spread function estimation," *IEEE Trans. Consum. Electron.*, vol. 56, no. 3, pp. 1204–1210, Aug. 2010.
[26] S. C. Hsu, S. F. Liang, K. W. Fan, and C. T. Lin, "A robust in-car digital image stabilization technique," *IEEE Trans. Syst., Man Cybern. C, Appl. Rev.*, vol. 37, no. 2, pp. 234–247, Feb. 2007.
[27] A. Sarhan and M. Faheem, "A proposed intelligent denoising technique for spatial video denoising for real-time applications," *Int. J. Mobile Computing Multimedia Commun.*, pp. 20–39, 2010.
[28] V. Murino, G. L. Foresti, and C. S. Regazzoni, "Adaptive camera regulation for investigation of real scenes," *IEEE Trans. Ind. Electron.*, vol. 43, no. 5, pp. 588–600, May 1996.
[29] B. Abidi, N. Aragam, Y. Yi, and M. Abidi, "Survey and analysis of multimodal sensor planning and integration for wide area surveillance," *ACM Computing Surveys*, vol. 41, no. 1, pp. 1–36, 2008.
[30] A. Elouardi, S. Bouaziz, A. Dupret, L. Lacassagne, J. O. Klein, and R. Reynaud, "A smart sensor-based vision system: Implementation and evaluation," *J. Phys. D, Appl. Phys.*, vol. 39, no. 8, pp. 1694–1705, 2006.
[31] S. Ghiasi, H. J. Moon, A. Nahapetian, and M. Sarrafzadeh, "Collaborative and reconfigurable object tracking," *J. Supercomputing*, vol. 30, no. 3, pp. 213–238, 2004.
[32] G. S. K. Fung, N. H. C. Yung, and G. K. H. Pang, "Camera calibration from road lane markings," *Opt. Eng.*, vol. 42, no. 10, pp. 2967–2977, 2003.
[33] S. C. Hui and F. Wang, "Remote video monitoring over the WWW," *Multimedia Tools Applic.*, vol. 21, no. 2, pp. 173–195, 2003.
[34] A. Badii, C. Lallah, M. Zhu, and M. Crouch, "Semi-automatic knowledge extraction, representation and context-sensitive intelligent retrieval of video content using collateral context modelling with scalable ontological networks," *Signal Process.-Image Commun.*, vol. 24, no. 9, pp. 759–773, 2009.
[35] G. Y. Hong, B. Fong, and A. C. M. Fong, "An intelligent video categorization engine," *Kybernetes*, vol. 34, no. 5–6, pp. 784–802, 2005.
[36] T.-H. Tsai, C.-L. Fang, and R.-C. Kuo, "A videotext in picture display system with embedded videotext extraction on a dual-core platform," *IEEE Trans. Consum. Electron.*, vol. 56, no. 3, pp. 1661–1669, Aug. 2010.
[37] T. Yadav and R. S. Aygun, "I-quest: An intelligent query structuring based on user browsing feedback for semantic retrieval of video data," *Multimedia Tools Applic.*, vol. 43, no. 2, pp. 145–178, 2009.
[38] Y. B. Kim and M. Shibata, "Content-based video indexing and retrieval—A natural language approach," *IEICE Trans. Inf. Syst.*, vol. E79D, no. 6, pp. 695–705, 1996.
[39] P. Shivakumara, R. Sreedhar, Q. P. Trung, L. Shijian, and L. T. Chew, "Multioriented video scene text detection through bayesian classification and boundary growing," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 8, pp. 1227–1235, 2012.
[40] P. Shivakumara, Q. P. Trung, and L. T. Chew, "A Laplacian approach to multi-oriented text detection in video," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 2, pp. 412–419, Feb. 2011.
[41] T. Mei, J. L. Guo, X. S. Hua, and F. L. Liu, "AdOn: Toward contextual overlay in-video advertising," *Multimedia Syst.*, vol. 16, no. 4–5, pp. 335–344, 2010.
[42] I. Grgurevic, A. Stancic, and P. Skorput, "Credibility and authenticity of digitally signed videos in traffic," *Promet-Traffic Transportation*, vol. 20, no. 6, pp. 405–414, 2008.
[43] P. Kumar, A. Mittal, and P. Kumar, "Study of robust and intelligent surveillance in visible and multimodal framework," *Informatica*, pp. 447–461, 2007.
[44] S. Kwak, G. Bae, and H. Byun, "Abandoned luggage detection using a finite state automaton in surveillance video," *Opt. Eng.*, vol. 49, no. 2, pp. 027007–1–02007–10, 2010.
[45] B. Krausz and R. Herpers, "MetroSurv: Detecting events in subway stations," *Multimedia Tools Applic.*, vol. 50, no. 1, pp. 123–147, 2010.
[46] M. Prieto and A. Allen, "Using self-organising maps in the detection and recognition of road signs," *Image Vis. Computing*, pp. 673–683, 2009.
[47] Y. C. Tsai, P. Kim, and Z. H. Wang, "Generalized traffic sign detection model for developing a sign inventory," *J. Computing Civil Eng.*, vol. 23, no. 5, pp. 266–276, 2009.
[48] J. L. Shih, Y. N. Chen, K. C. Yan, and C. C. Han, "Illegal entrant detection at a restricted area in open spaces using color features," *J. Inf. Sci. Eng.*, vol. 25, no. 5, pp. 1575–1592, 2009.
[49] C. Stauffer and W. Grimson, "Adaptive background mixture models for real-time tracking," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 1999, pp. 246–252.
[50] T. Gao, Z. G. Liu, S. H. Yue, J. Q. Mei, and J. Zhang, "Traffic video-based moving vehicle detection and tracking in the complex environment," *Cybern. Syst.*, vol. 40, no. 7, pp. 569–588, 2009.

[51] C.-T. Li, C.-T. Yang, S. Yu-Wen, and S. Tzu-Kuei, “An efficient and robust moving shadow removal algorithm and its applications in ITS,” *EURASIP J. Adv. Signal Process.*, vol. 2010, 2010, Art. ID 945130.

[52] D.-M. Tsai, Y. Chiang, and Y.-H. Tsai, “A shift-tolerant dissimilarity measure for surface defect detection,” *IEEE Trans. Ind. Inf.*, vol. 8, no. 1, pp. 128–137, Feb. 2012.

[53] S. Chen, J. Zhang, Y. Li, and J. Zhang, “A hierarchical model incorporating segmented regions and pixel descriptors for video background subtraction,” *IEEE Trans. Ind. Inf.*, vol. 8, no. 1, pp. 118–127, Feb. 2012.

[54] W. Choi, C. Pantofaru, and S. Savarese, “A general framework for tracking multiple people from a moving camera,” *IEEE Trans. Pattern Anal. Mach. Intell.*, to be published.

[55] T. Wang, Z. Zhu, and E. Blasch, “Bio-inspired adaptive hyperspectral imaging for real-time target tracking,” *IEEE Sensors J.*, vol. 10, no. 3, pp. 647–654, Mar. 2010.

[56] T. Gandhi and M. M. Trivedi, “Vehicle surround capture: Survey of techniques and a novel omni-video-based approach for dynamic panoramic surround maps,” *IEEE Trans. Intell. Transp. Syst.*, vol. 7, no. 3, pp. 293–308, Sep. 2006.

[57] C. Liu and N. Yung, “Scale-adaptive spatial appearance feature density approximation for object tracking,” *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 1, pp. 284–290, Mar. 2011.

[58] W. L. Hsu, C. C. Hsiao, Y. L. Chang, and T. L. Chen, “Vision-based monitoring method using gray relational analysis,” *IET Comput. Vis.*, vol. 3, no. 3, pp. 103–111, 2009.

[59] Y. Tie and L. Guan, “Automatic face detection in video sequences using local normalization and optimal adaptive correlation techniques,” *Pattern Recognit.*, vol. 42, no. 9, pp. 1859–1868, 2009.

[60] L. Li, H. Weimin, I. Gu, L. Ruijiang, and T. Qi, “An efficient sequential approach to tracking multiple objects through crowds for real-time intelligent CCTV systems,” *IEEE Trans. Syst., Man Cybern., Cybern.*, vol. 38, no. 5, pp. 1254–1269, 2008.

[61] X. Zhang, S. Hu, D. Chen, and X. Li, “Fast covariance matching with fuzzy genetic algorithm,” *IEEE Trans. Ind. Inf.*, vol. 8, no. 1, pp. 148–157, Feb. 2012.

[62] G. Diamantopoulos and M. Spann, “Event detection for intelligent car park video surveillance,” *Real-Time Imaging*, vol. 11, no. 3, pp. 233–243, 2005.

[63] K. Q. Huang and T. N. Tan, “Vs-star: A visual interpretation system for visual surveillance,” *Pattern Recognit. Lett.*, vol. 31, no. 14, pp. 2265–2285, 2010.

[64] W. Zhang, Q. M. J. Wu, and H. B. Yin, “Moving vehicles detection based on adaptive motion histogram,” *Digit. Signal Process.*, vol. 20, no. 3, pp. 793–805, 2010.

[65] Z. N. Zhou, X. Chen, Y. C. Chung, Z. H. He, T. X. Han, and J. M. Keller, “Activity analysis, summarization, visualization for indoor human activity monitoring,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 11, pp. 1489–1498, Nov. 2008.

[66] L. Qu and Y. Chen, “Frequent trajectory patterns mining for intelligent visual surveillance system,” *J. Donghua Univ.*, pp. 164–170, 2009.

[67] H. Veeraraghavan and N. P. Papanikolopoulos, “Learning to recognize video-based spatiotemporal events,” *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 4, pp. 628–638, Dec. 2009.

[68] J. Yuen and E. Lee, “The effect of overtaking behavior on unidirectional pedestrian flow,” *Safety Sci.*, vol. 50, no. 8, pp. 1704–1714, 2012.

[69] L. Atallah and G. Z. Yang, “The use of pervasive sensing for behaviour profiling—A survey,” *Pervasive and Mobile Computing*, vol. 5, no. 5, pp. 447–464, 2009.

[70] T. K. Ho, K. Matthews, L. O’Gorman, and H. Steck, “Public space behavior modeling with video and sensor analytics,” *Bell Labs Tech. J.*, vol. 16, no. 4, pp. 203–217, 2012.

[71] J. Albusac, D. Vallejo, L. Jimenez-Linares, J. J. Castro-Schez, and L. Rodriguez-Benitez, “Intelligent surveillance based on normality analysis to detect abnormal behaviors,” *Int. J. Pattern Recognit. Artif. Intell.*, vol. 23, no. 7, pp. 1223–1244, 2009.

[72] E. Fellores, “Chilean salmon farm enhances security with intelligent video edge,” *Security*, p. 22, 2008.

[73] R. Chang, L. Guan, and J. A. Burne, “An automated form of video image analysis applied to classification of movement disorders,” *Disabil. Rehabil.*, vol. 22, no. 1–2, pp. 97–108, 2000.

[74] T. Lee and A. Mihailidis, “An intelligent emergency response system: Preliminary development and testing of automated fall detection,” *J. Telemed. Telecare*, vol. 11, no. 4, pp. 194–198, 2005.

[75] S. T. Londei, J. Rousseau, F. Ducharme, A. St-Arnaud, J. Meunier, J. Saint-Arnaud, and F. Giroux, “An intelligent videomonitoring system for fall detection at home: Perceptions of elderly people,” *J. Telemed. Telecare*, vol. 15, no. 8, pp. 383–390, 2009.

[76] H. Rastgar, M. Ahmadi, and M. Sid-Ahmed, “A stereo vision-based bin picking system using Hopfield neural networks,” *J. Circuits Syst. Comput.*, vol. 18, no. 3, pp. 443–463, 2009.

[77] J. X. Wang, G. Bebis, M. Nicolescu, M. Nicolescu, and R. Miller, “Improving target detection by coupling it with tracking,” *Mach. Vis. Applications*, vol. 20, no. 4, pp. 205–223, 2009.

[78] S. Y. Chen, “Kalman filter for robot vision: A Survey,” *IEEE Trans. Ind. Electron.*, vol. 59, no. 11, pp. 4409–4420, Nov. 2012.

[79] J. Chen, E. Tan, and Z. D. Li, “A machine learning framework for real-time traffic density detection,” *Int. J. Pattern Recognit. Artif. Intell.*, vol. 23, no. 7, pp. 1265–1284, 2009.

[80] S. Kumarawadu, K. Watanabe, and T. T. Lee, “High-performance object tracking and fixation with an online neural estimator,” *IEEE Trans. Syst., Man Cybern. B, Cybern.*, vol. 37, no. 1, pp. 213–223, Mar. 2007.

[81] P. J. Wang, C. M. Li, C. Y. Wu, and H. J. Li, “A channel awareness vehicle detector,” *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 2, pp. 339–347, Jun. 2010.

[82] B. Takacs, “Real-time millimeter wave radar vision system,” *J. Imaging Sci. Technol.*, vol. 49, no. 3, pp. 230–236, 2005.

[83] M. Kafai and B. Bhanu, “Dynamic bayesian networks for vehicle classification in video,” *IEEE Trans. Ind. Inf.*, vol. 8, no. 1, pp. 100–109, Feb. 2012.

[84] F. Tung, J. S. Zelek, and D. A. Clausi, “Goal-based trajectory analysis for unusual behaviour detection in intelligent surveillance,” *Image Vis. Computing*, vol. 29, no. 4, pp. 230–240, 2011.

[85] C. L. Chen, X. Tao, and G. Shaogang, “Detecting and discriminating behavioural anomalies,” *Pattern Recognit.*, vol. 44, no. 1, pp. 117–132, 2011.

[86] T. Michalke, J. Fritsch, and C. Goerick, “A biologically-inspired vision architecture for resource-constrained intelligent vehicles,” *Comput. Vis. Image Understanding*, vol. 114, no. 5, pp. 548–563, 2010.

[87] C. Quek, W. Irawan, and E. Ng, “A novel brain-inspired neural cognitive approach to SARS thermal image analysis,” *Expert Syst. Applic.*, pp. 3040–3054, 2010.

[88] A. M. Callan, R. Osu, Y. Yamagishi, D. E. Callan, and N. Inoue, “Neural correlates of resolving uncertainty in driver’s decision making,” *Human Brain Mapping*, vol. 30, no. 9, pp. 2804–2812, 2009.

[89] M. Nielsen, D. C. Slaughter, and C. Gliever, “Vision-based 3D peach tree reconstruction for automated blossom thinning,” *IEEE Trans. Ind. Inf.*, vol. 8, no. 1, pp. 188–196, Feb. 2012.

[90] T. Cuong and M. M. Trivedi, “3-D posture and gesture recognition for interactivity in smart spaces,” *IEEE Trans. Ind. Inf.*, vol. 8, no. 1, pp. 178–187, Feb. 2012.

[91] M. K. Habib, “Fiber-grating-based vision system for real-time tracking, monitoring, obstacle detection,” *IEEE Sensors J.*, vol. 7, no. 1–2, pp. 105–121, 2007.

[92] D. M. Gavrila and S. Munder, “Multi-cue pedestrian detection and tracking from a moving vehicle,” *Int. J. Comput. Vis.*, vol. 73, no. 1, pp. 41–59, 2007.

[93] N. Pavlidou, N. Grammalidis, K. Dimitropoulos, D. Simitopoulos, M. Strintzis, A. Gilbert, E. Piazza, C. Herrlich, and R. Heidger, “Using intelligent digital cameras to monitor aerodrome surface traffic,” *IEEE Intell. Syst.*, vol. 20, no. 3, pp. 76–81, 2005.

[94] J. Cai, M. Shehata, and W. Badawy, “A robust video-based algorithm for detecting snow movement in traffic scenes,” *J. Signal Process. Syst. Signal Image Video Technol.*, vol. 56, no. 2–3, pp. 307–326, 2009.

[95] C. L. Mak and H. S. L. Fan, “Heavy flow-based incident detection algorithm using information from two adjacent detector stations,” *J. Intell. Transp. Syst.*, vol. 10, no. 1, pp. 23–31, 2006.

[96] K. Deb and K. H. Jo, “A vehicle license plate detection method for intelligent transportation system applications,” *Cybern. Syst.*, vol. 40, no. 8, pp. 689–705, 2009.

[97] S. Gidel, P. Checchin, C. Blanc, T. Chateau, and L. Trassoudaine, “Pedestrian detection and tracking in an urban environment using a multilayer laser scanner,” *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 3, pp. 579–588, Sep. 2010.

[98] S. J. Krotosky and M. M. Trivedi, “On color-, infrared-, multimodal-stereo approaches to pedestrian detection,” *IEEE Trans. Intell. Transp. Syst.*, vol. 8, no. 4, pp. 619–629, Dec. 2007.

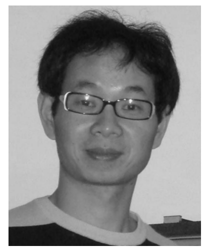
[99] Y. C. Tsai, Z. Z. Hu, and C. Alberti, “Detection of roadway sign condition changes using multi-scale sign image matching (M-SIM),” *Photogramm. Eng. Remote Sens.*, vol. 76, no. 4, pp. 391–405, 2010.

[100] W. Wu, X. L. Chen, and J. Yang, “Detection of text on road signs from video,” *IEEE Trans. Intelligent Transportation Systems*, vol. 6, no. 4, pp. 378–390, 2005.
[101] P. C. Su and C. Y. Wu, “A joint watermarking and ROI coding scheme for annotating traffic surveillance videos,” *Eurasip Journal on Advances in Signal Processing*, 2010.
[102] K. Dimitropoulos, N. Grammalidis, D. Simitopoulos, N. Pavlidou, and M. Strintzis, “Video system for surface movement surveillance at airports,” *Journal of Intelligent Transportation Systems*, vol. 11, no. 4, pp. 169–180, 2007.
[103] R. Nagarajan, S. Yaacob, and G. Sainarayanan, “Computer aided vision assistance for human blind,” *Integr. Comput.-Aided Eng.*, vol. 11, no. 1, pp. 15–24, 2004.
[104] C. Connolly, “Wildlife-spotting robots,” *Sensor Rev.*, vol. 27, no. 4, pp. 282–287, 2007.
[105] A. Baca, P. Dabnichki, M. Heller, and P. Kornfeind, “Ubiquitous computing in sports: A review and analysis,” *J. Sports Sci.*, vol. 27, no. 12, pp. 1335–1346, 2009.
[106] L. Snidaro, C. Micheloni, and C. Chiavedale, “Video security for ambient intelligence,” *IEEE Trans. Syst., Man Cybern. A, Syst. Humans*, vol. 35, no. 1, pp. 133–144, Jan. 2005.
[107] P. Leed, A. Fulton, K. Raychaudhuri, and A. Esquivel, “Preemptive response to missile threat using intelligent video,” *J. Aerosp. Computing Inf. Commun.*, vol. 7, no. 6, pp. 188–195, 2010.
[108] J. Vanne, E. Aho, K. Kuusilinna, and T. Hamalainen, “A configurable motion estimation architecture for block-matching algorithms,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 4, pp. 466–477, Apr. 2009.
[109] T. Kohler, F. Rochter, J. Lindemann, and R. Moller, “Bio-inspired motion detection in an FPGA-based smart camera module,” *Bioinspiration Biomimetics*, 2009, Art. ID 015008.
[110] Y. P. Huang, C. L. Chiou, and F. E. Sandnes, “An intelligent strategy for the automatic detection of highlights in tennis video recordings,” *Expert Syst. Applic.*, vol. 36, no. 6, pp. 9907–9918, 2009.
[111] M. S. Hossain, M. M. Hassan, M. A. Qurishi, and A. Alghamdi, “Resource allocation for service composition in cloud-based video surveillance platform,” in *Proc. IEEE Int. Conf. Multimedia Expo Workshops*, 2012, pp. 408–412.
[112] D. A. Rodriguez-Silva, L. dkinson-Orellana, F. J. Gonz’lez-Castano, I. rmino-Franco, and D. Gonz’lez-Martinez, “Video surveillance based on cloud storage,” in *Proc. IEEE 5th Int. Conf. Cloud Computing*, 2012, pp. 991–992.
[113] S. T. Shivappa, M. M. Trivedi, and B. D. Rao, “Audiovisual information fusion in human-computer interfaces and intelligent environments: A survey,” *Proc. IEEE*, vol. 98, no. 10, pp. 1692–1715, Oct. 2010.
[114] R. I. Chang, T. C. Wang, C. H. Wang, J. C. Liu, and J. M. Ho, “Effective distributed service architecture for ubiquitous video surveillance,” *Inf. Syst. Frontiers*, vol. 14, no. 3, pp. 499–515, 2012.

Honghai Liu (M’02–SM’06) received the Ph.D. degree in robotics from King’s College London, London, U.K., in 2003.

He joined the University of Portsmouth, Portsmouth, U.K., in September 2005. He previously held research appointments at the Universities of London and Aberdeen and project leader appointments in large-scale industrial control and system integration industry. He has authored and coauthored over 250 peer-reviewed international journals and conference papers. He is an associate editor for the *International Journal of Fuzzy Systems.*He is interested in approximate computation, pattern recognition, intelligent video analytics, and cognitive robotics and their practical applications with an emphasis on approaches which could make contribution to the intelligent connection of perception to action using contextual information.

Dr. Liu was the recipient of four Best Paper Awards. He is an associate editor for the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, the IEEE TRANSACTIONS ON HUMAN–MACHINES SYSTEMS, and the IEEE TRANSACTIONS ON FUZZY SYSTEMS.



Shengyong Chen (M’01–SM’10) received the Ph.D. degree in computer vision from City University of Hong Kong, Hong Kong, in 2003.

He joined Zhejiang University of Technology, China, in February 2004, where he is currently a Professor with the Department of Computer Science. He received a fellowship from the Alexander von Humboldt Foundation of Germany and was with the University of Hamburg, Hamburg, Germany, in 2006–2007. He was a Visiting Professor with Imperial College, London, U.K., in 2008–2009 and a Visiting Professor with the University of Cambridge, Cambridge, U.K., in 2012. His research interests include computer vision, robotics, 3-D object modeling, and image analysis. He has published over 100 scientific papers in international journals and conferences.

Dr. Chen is a Fellow of IET and a committee member of IET Shanghai Branch. He was the recipient of the Best Paper Award of the IEEE SENSORS JOURNAL in 2012.

Naoyuki Kubota (S’95–A’97–M’01) received the B.Sc. degree from Osaka Kyoiku University, Kashiwara, Japan, in 1992, the M.Eng. degree from Hokkaido University, Hokkaido, Japan, in 1994, and the D.E. degree from Nagoya University, Nagoya, Japan, in 1997.

He joined the Osaka Institute of Technology, Osaka, Japan, in 1997. In 2000, he joined the Department of Human and Artificial Intelligence Systems, Fukui University, as an Associate Professor. He joined the Department of Mechanical Engineering, Tokyo Metropolitan University, in 2004, and has been an Associate Professor with the Department of System Design, Tokyo Metropolitan University, Tokyo, Japan, since 2005.