Special Issue on Video Surveillance

VIDEO surveillance is a rather broad concept. Apart from safety and security, video surveillance has a wide variety of applications in numerous other aspects of life. Digital video pays a pivotal role in a myriad of surveillance applications. Automated means of video surveillance for public safety enhancement have existed for quite some time but have gained significant popularity only in recent years. Enhanced awareness for public safety is leading to innovative research by making use of multimedia information and telecommunications for disaster and crime prevention and management of secure environments. A new generation of video surveillance is emerging with innovative functionalities aided by new scientific rigor in areas such as communication, compression, data mining, content-based video retrieval, machine learning, and pattern recognition. Tracking humans, objects, and motion, for instance, can provide several assistive means in social environments, such as helping disabled people and enhancing manufacturing productivity. Identifying, tracking, and monitoring activities can lead to behavior analysis and understanding. All of this, in turn, is leading to new applications of surveillance in homeland security and crime prevention through indoor and outdoor monitoring and monitoring of critical infrastructures, highways, parking garages, and shopping malls.

As such, several research challenges arise in system design for collection and dissemination of video data as well as algorithms for processing the data collected, for its meaningful interpretation in the ongoing context and for carrying out automated security services accordingly. This Special Issue seeks to consolidate the broader range of the recent research achievements in developing new theories, algorithms, architectures, and integrated video platforms that facilitate or enable video surveillance.

This Special Issue received 66 paper submissions, posing a major challenge for reviewing. We sought assistance for reviewing these papers from a large number of reviewers. Each paper was reviewed by three to five reviewers, ensuring high reviewing standards. All papers were also revised to improve quality. Since the quality of most submitted papers was very high, we had to reject several good papers. Based on the reviewers' recommendations and after considerable deliberation, we selected the following 14 regular and two brief papers. These papers capture some of the state-of-the-art research on video surveillance issues, provide comprehensive overview of existing techniques, and propose novel solutions for important research problems. We hope the special issue will prove to be an invaluable resource for researchers, engineers, and practitioners engaged in video surveillance.

In their paper, "High-speed action recognition and localization in compressed domain videos," Yeo *et al.* consider the problem of action recognition and localization from a video sequence. They propose a technique that aims to detect occurrences of an action in a test video, at specific times and locations. This has a direct application in video surveillance, where it is critical to be able to respond to events as they happen. One challenge is to perform such action recognition in real time. Their technique works quickly because it takes advantage of the fact that some of the video processing for compression can be reused in video analysis or transcoding. Their technique assumes that a surveillance application would consist of a front-end system that records, compresses, stores, and transmits videos as well as a back-end system that processes the transmitted video to accomplish various tasks. This architecture facilitates the execution of action recognition task at the back-end but allows other jobs, such as the choice of video coding method, to be made at the front-end. The paper discusses how various video coding settings impact the action recognition performance of their approach. They present results on the compression classification tradeoff, which would provide valuable insight into jointly designing a system that performs video encoding at the camera front-end and action classification at the processing back-end.

In "Efficient multi-target visual tracking using random finite sets," Maggio *et al.* propose a probability hypothesis density (PHD) filter-based multitarget tracking scheme. The PHD filter is able to compensate miss detections and remove noise and clutter that are very commonly encountered in a tracking process. The authors propose to simplify the multitarget tracking process by propagating the first-order moment of the multitarget posterior. To implement the propagation scheme, they use a particle resampling strategy to cope with the varying object scales. The resampling strategy enables the use of the PHD filter when *a priori* knowledge of the scene is not available. The designed dynamic and observation models can be applied to any Bayesian tracker. Experiments reported in the paper demonstrate the effectiveness of the proposed scheme.

Camera motion parameter estimation is an important topic in video surveillance. It is the enabling step for moving object detection and tracking and further video content analysis tasks. In "Camera motion estimation using a novel online vector field model in particle filters," Nikitidis *et al.* propose a new scheme for parametric camera motion estimation based on a stochastic vector field model. Their technique handles both smooth and rapid camera motion changes. Embedded in a particle filter framework, the technique predicts the future camera motion based on current and prior observations. The authors also incorporate their proposed camera motion estimation scheme into an object tracker to demonstrate its efficiency in moving object tracking.

In "A unified framework for consistent 2-D/3-D foreground object detection," Landabaso and Pardas develop a Bayesian framework for 2-D and 3-D foreground segmentation. This Bayesian framework provides a unified manner to interact between the planar and the volumetric detection tasks and

Digital Object Identifier 10.1109/TCSVT.2008.929646

helps to prevent the propagation of noisy pixel observations to the 3-D space. They observe that errors in 2-D foreground detection often produce a set of incompatible foreground planar regions in the sense that they cannot be globally explained as the projection of the detected 3-D volume. To address this issue, they develop a new 3-D foreground detection scheme that is able to correct errors in 2-D planar detections by checking the consistency between 3-D foreground detections and the set of corresponding 2-D foreground regions.

In "Heterogeneous fusion of omnidirectional and PTZ cameras for multiple object tracking," Chen *et al.* propose a dual-camera system which can be used in a video surveillance system. Dual-camera systems have been widely adopted in surveillance because of their better coverage of potential field-of-view (omnidirectional camera) and wider zoom range of the PTZ camera. Most existing algorithms need *a priori* knowledge of the omnidirectional camera's projection model to solve the nonlinear spatial correspondence problem. The authors propose a new technique that uses geometry and homography calibration to approximate the camera's projection model and spatial mapping, respectively. The proposed methods improve not only the mapping accuracy but also the computation load and flexibility.

In "Probabilistic object tracking with dynamics attributed relational feature graph," Tang and Tao address object tracking, recognizing that its performance strongly depends on the effective object representation and the adequate updating of this representation. The authors propose a solution for tracking objects with changing appearance based on a novel sparse, local feature-based object representation, called an attributed relational feature graph (ARFG). In this context, the object is modeled using invariant features, such as the scale-invariant feature transform (SIFT), while the geometric relations among features are encoded using a graph. A dynamic model is proposed to evolve the feature graph depending on the appearance and structure changes by adding new stable features as well as removing inactive features. Experimental results show that the proposed tracker achieves robust tracking under critical conditions including significant appearance changes, view point changes, and occlusions.

Stereo-based pedestrian detection techniques are used for automatic pedestrian detection, counting, and tracking. However, due to the lack of standard test data and an agreed methodology for carrying out the evaluation, a quantitative assessment of the performance has been problematic. In "SIVS: A framework for evaluating stereo-based pedestrian detection techniques," Kelly *et al.* propose a framework for the quantitative evaluation of a stereo-based pedestrian detection system. The framework uses a number of publicly available test sequences and groundtruth sets that incorporate many of the challenges that are inherent for pedestrian detection in real application scenarios. They provide freely available synthetic and real-world test data and recommend a set of evaluation metrics. This allows researchers to benchmark systems.

Ni *et al.*, in their paper, "A hybrid framework for 3-D human motion tracking," address the important problem of multiview-based marker-less articulated human motion tracking problem. They propose a solution that is based on a hybrid framework for articulated 3-D human motion tracking from multiple synchronized cameras with potential uses in surveillance systems. In contrast to previous works that rely on deterministic search or stochastic sampling, they utilize a hybrid sample-and-refine framework that combines both stochastic sampling and deterministic optimization to achieve a good compromise between efficiency and robustness. Similar motion patterns are used to learn a compact low-dimensional representation of the motion statistics. Their technique implements sampling in a low-dimensional space during tracking, which reduces the number of particles drastically. For further improving the optimality of the tracking, they utilize a local optimization method based on simulated physical force/moment into their tracking framework,.

In "Data-driven probability hypothesis density filter for visual tracking," Wang *et al.* propose to apply the probability hypothesis density (PHD) filter to track an arbitrary number of pedestrians in image sequences. They use a particle filter to implement the PHD filter. It is known that designing the importance function for a particle filter-based filter is still a challenging issue, because the processed targets can sometimes appear, sometimes disappear, or even split and merge at any time. To tackle the above-mentioned challenges, the authors propose to model the targets into two categories: survival objects and spontaneous birth objects. Based on the proposed models, they derive a data-driven importance function for a particle PHD filter and then use it in the pedestrian tracking process.

Person analysis in multispectral and multiperspective imagery is a relatively new area of research in video surveillance. In "Person surveillance using visual and infrared imagery," Krotosky and Trivedi present a methodology for analyzing multimodal and multiperspective systems for person surveillance. Using an experimental testbed consisting of two color and two infrared cameras, they can accurately register the color and infrared imagery for any general scene configuration so the scope of multispectral analysis can be expanded beyond the specialized long-range surveillance experiments of previous approaches to more general scene configurations common to unimodal approaches.

Jacobs and Pless take an interesting approach to understanding the temporal structure of video surveillance data. In their paper, "Time scales in video surveillance," they propose a set of filters to define a temporal scale-space representation for the activity at each pixel. The advantage of this approach is that scale space can be maintained and continuously updated in real time. This allows characterizing of interesting temporal features and facilitates approximate reconstruction of the video history under challenging noise conditions. In contrast to spatial scale spaces, the values of the pixel measured by filters with varying temporal scales are shown to be useful as a feature defining the recent history of that pixel. The paper demonstrates how these features correlate with scene properties of interest.

Vision-based trajectory learning and analysis is an important research problem. In their paper, "A survey of vision-based trajectory learning and analysis for surveillance," Morris and Trivedi provide a comprehensive overview of several existing techniques the have been proposed in the literature. They focus on various real-time techniques that use trajectory data to define a general set of activities that are applicable to a wide range of

scenes and environments. Typically, the events of interest are detected by building a generic topological scene description from an underlying motion structure observed over time. The scene topology is distinguished by points of interest (POIs) and characterizes activity paths (PAs) by the way objects move between POIs. The paper presents in detail the methods that automatically learn the POI/AP model based on data. The paper indexes these methods based on several criteria, such as the type of input for the learning algorithms, the choice of clustering scheme, and the variety of analyses provided. In these methods, the scene description is learnt in an unsupervised fashion to accurately portray the data and is sufficient to define a diverse set of events for further analysis triggering, including virtual fencing, speed profiling, behavior classification, anomaly detection, and object interaction.

In "Activity recognition using a combination of category components and local models for video surveillance," Lin *et al.* present a new approach for automatic recognition of human activities for video surveillance applications. Specifically, they represent an activity by a combination of category components. This approach offers flexibility to add new activities to the system and an ability to deal with the problem of building models for activities lacking training data. For improving the recognition accuracy, they also propose a confident-frame-based recognition algorithm, where the video frames with high confidence for recognizing an activity are used as a specialized local model to help classify the remainder of the video frames. They present an efficient algorithm to improve the accuracy of recognition.

Because human beings are typically an important surveillance target, detecting humans, notably pedestrians, is a very relevant task which has to be accomplished with a good performance. In "Fast pedestrian detection using a cascade of boosted covariance features," Paisitkriangkrai *et al.* propose a solution for fast pedestrian detection based on covariance features. These features have been determined based on a comprehensive experimental study on pedestrian detection using state-of-the-art locally extracted features. For complexity reasons, the features are selected and the classifiers trained in a Euclidean space. A cascaded classifier structure based on AdaBoost with weighted Fisher linear discriminant analysis-based weak classifiers is constructed for efficient detection. Experimental results show the proposed detector provides good detection performance with lower complexity. This complexity performance is further improved by using a multiple layer boosting with heterogeneous features to exploit the efficiency of the Haar-like feature and the discriminative power of the covariance feature.

With the proliferation of video surveillance, people are increasingly concerned about the invasiveness of ubiquitous surveillance and fear that their privacy is at risk. On the other hand, law enforcement agencies aim to prevent and prosecute criminal activities. Consequently, the need for private organizations to protect against unauthorized activities on their premises is often in conflict with the privacy requirements of individuals. Martin and Plataniotis, in their paper, "Privacy protected surveillance using secure visual object coding," present a shape and texture coding scheme that can be employed in privacy-protected surveillance systems. In this scheme, visual objects are encrypted so that the content is only available to authorized personnel with the correct decryption key. The secure visual object coder employs a shape and texture partitioning scheme combined with a selective encryption scheme for efficient, secure storage and transmission of visual object shapes and textures. Since the encryption is performed in the compressed domain, it does not affect the rate-distortion performance of the coder. The paper proposes the usage of a separate parameter for each encrypted object to control the strength of the encryption versus required processing overhead. The paper provides security analyses, demonstrating the confidentiality of both the encrypted and unencrypted portions of the secured output bit-stream.

Finally, the last paper also addresses privacy protection in video surveillance. In "Scrambling for privacy protection in video surveillance systems," Dufaux and Ebrahimi introduce several approaches to conceal regions of interest based on transform-domain scrambling in the MPEG-4 compressed video. The approaches pseudorandomly flip the sign of selected transform coefficients, or invert some bits of the codestream. The simulation results presented in the paper show the proposed techniques conceal privacy-sensitive information while the scene remains comprehensible. The impact of these approaches on coding efficiency is small and the required computational complexity is negligible. Moreover, the methods are shown to be secure against brute-force or error concealment attacks.

We would like to thank the Editor-in-Chief, Dr. Chang Wen Chen, for his constant support and useful advice as well as the rest of the Editorial Board for accepting this Special Issue among a number of competing Special Issue proposals. We are also thankful to all of the reviewers who took time out of their busy schedules to perform the review service. These detailed reviews have been instrumental in selecting the final set of papers and in helping the authors to improve the technical and presentation quality of their papers. Last but not least, we would like to thank all of the authors who submitted papers for this Special Issue.

ISHFAQ AHMAD, *Guest Editor*
University of Texas at Arlington
Arlington, TX 76109 USA

ZHIHAI HE, *Guest Editor*
University of Missouri
Columbia, MO 65211 USA

MARK LIAO, *Guest Editor*
Academia Sinica
11529 Taipei, Taiwan, R.O.C.

FERNANDO PEREIRA, *Guest Editor*
Instituto Superior Tecnico
Lisbon, Portugal

MING-TING SUN, *Guest Editor*
University of Washington
Seattle, WA 98195 USA

Ishfaq Ahmad (F'08) received the B.Sc. degree in electrical engineering from the University of Engineering and Technology, Lahore, Pakistan, in 1985 and the M.S. degree in computer engineering and the Ph.D. degree in computer science from Syracuse University, Syracuse, NY, in 1987 and 1992, respectively.

He is currently a Professor of computer science and engineering with the University of Texas at Arlington (UTA). Prior to joining UTA, he was an Associate Professor with the Computer Science Department, Hong Kong University of Science and Technology. At UTA, he leads the Multimedia Laboratory and the Institute for Research in Security (IRIS). IRIS, an interdisciplinary research center spanning several departments, is engaged in research on advanced technologies for homeland security and law enforcement. He is known for his research contributions in parallel and distributed computing, multimedia computing, video compression, and security. His work in these areas has been published in approximately 200 technical papers in peer-reviewed journals and conferences. He has worked extensively with industry and delivered several techniques to both start-up and established companies.

Prof. Ahmad was a recipient of numerous awards, which include three Best Paper Awards at leading conferences, the 2007 Best Paper Award for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, and the IEEE Service Appreciation Award. His current research is funded by the Department of Justice, the National Science Foundation, and industry. He is an Associate Editor of the *Journal of Parallel and Distributed Computing,* the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, and the IEEE TRANSACTIONS ON MULTIMEDIA. He is the Guest Editor for a Special Issue of the IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS on Power-Aware Parallel and Distributed Systems (PAPADS).

Zhihai He received the B.S. degree from Beijing Normal University, Beijing, China, in 1994 and the M.S. degree from the Institute of Computational Mathematics, Chinese Academy of Sciences, Beijing, China, in 1997, both in mathematics, and the Ph.D. degree in electrical engineering from the University of California, Santa Barbara, in 2001.

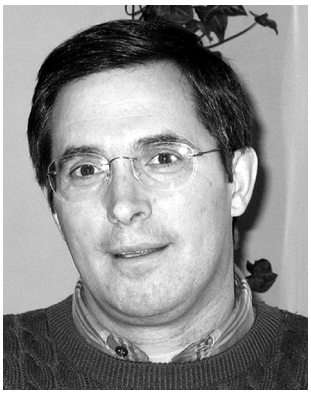
In 2001, he joined Sarnoff Corporation, Princeton, NJ, as a Member of Technical Staff. In 2003, he joined the Department of Electrical and Computer Engineering, University of Missouri, Columbia, as an Assistant Professor. His current research interests include image/video processing and compression, network transmission, wireless communication, computer vision analysis, sensor networks, and embedded system design.

Dr. He was the recipient of the 2002 IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY Best Paper Award and the SPIE VCIP Young Investigator Award in 2004. Currently, he serves as an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY and the *Journal of Visual Communication and Image Representation.* He is a member of the Visual Signal Processing and Communication Technical Committee of the IEEE Circuits and Systems Society and serves as Technical Program Committee member or session chair of a number of international conferences.

Mark Liao received the B.S. degree in physics from National Tsing-Hua University, Hsin-Chu, Taiwan, R.O.C., in 1981 and the M.S. and Ph.D. degrees in electrical engineering from Northwestern University, Evanston, IL, in 1985 and 1990, respectively.

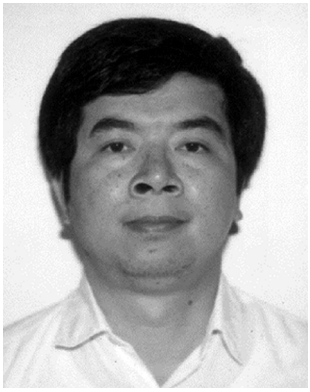
He was a Research Associate with the Computer Vision and Image Processing Laboratory, Northwestern University, during 1990–1991. In July 1991, he joined the Institute of Information Science, Academia Sinica, Taipei, Taiwan, R.O.C., as an Assistant Research Fellow. He was promoted to Associate Research Fellow and then Research Fellow in 1995 and 1998, respectively. From August 1997 to July 2000, he served as the Deputy Director of the institute. From February 2001 to January 2004, he served as the Acting Director of the Institute of Applied Science and Engineering Research. He is also jointly appointed as a Professor with the Computer Science and Information Engineering Department, National Chiao Tung University. His current research interests include multimedia signal processing, video-based surveillance systems, content-based multimedia retrieval, and multimedia protection. He is now the Editor-in-chief of the *Journal of Information Science and Engineering*. He is on the editorial boards of the *International Journal of Visual Communication and Image Representation*, *EURASIP Journal on Advances in Signal Processing,* and the *Research Letters in Signal Processing.*

Dr. Liao was the recipient of the Young Investigators' award from Academia Sinica in 1998 and the Excellent Paper Award from the Image Processing and Pattern Recognition society of Taiwan in 1998 and 2000. He was also the recipient of the Distinguished Research Award from the National Science Council of Taiwan in 2003 and the National Invention Award of Taiwan in 2004. He served as the Program co-chair of the Second IEEE Pacific-Rim conference on Multimedia 2001). In June 2004, he served as the conference co-chair of the 5th International Conference on Multimedia and Exposition (ICME 2004) and technical co-chair of 2007 ICME. He also served as a committee member of 2005, 2006, and 2007 ACM Multimedia Conference and 2007 World Wide Web Conference. He was an Associate Editor of the IEEE TRANSACTIONS ON MULTIMEDIA during 1998–2001.



Fernando Pereira (F'08) is currently a Professor with the Electrical and Computers Engineering Department, Instituto Superior Técnico (IST), Lisboa, Portugal. He is responsible for the participation of IST in many national and international research projects. He acts often as project evaluator and auditor for various organizations. He is an Area Editor of the *Signal Processing: Image Communication Journal*. His areas of interest are video analysis, processing, coding and description, and interactive multimedia services.

Dr. Pereira is or has been an Associate Editor of the IEEE TRANSACTIONS OF CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON MULTIMEDIA, and the IEEE SIGNAL PROCESSING MAGAZINE. He is a Member of the IEEE Signal Processing Society Image and Multiple Dimensional Signal Processing Technical Committee and of the IEEE Signal Processing Society Multimedia Signal Processing Technical Committee. He was an IEEE Distinguished Lecturer in 2005. He has been a member of the Scientific and Program Committees of many international conferences and has contributed more than 200 papers. He has been participating in the work of ISO/MPEG for many years, notably as the head of the Portuguese delegation, Chairman of the MPEG Requirements Group, and chairing many Ad Hoc Groups related to the MPEG-4 and MPEG-7 standards.



Ming-Ting Sun (S'79–M'81–SM'89–F'96) received the B.S. degree from National Taiwan University, Taipei, Taiwan, R.O.C., in 1976, the M.S. degree from the University of Texas at Arlington in 1981, and the Ph.D. degree from the University of California, Los Angeles, in 1985, all in electrical engineering.

He joined the University of Washington, Seattle, in August 1996, where he is a Professor. Previously, he was the Director of the Video Signal Processing Research Group with Bellcore. He was a Chaired Professor with TsingHwa University, Beijing, China, and a Visiting Professor with Tokyo University, Tokyo, Japan, and National Taiwan University. He holds 11 patents and has published over 200 technical papers, including 13 book chapters in the area of video and multimedia technologies. He coedited a book, *Compressed Video Over Networks*.

Dr. Sun was the recipient of an IEEE Circuits and Systems (CAS) Society Golden Jubilee Medal in 2000, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY (TCSVT) Best Paper Award in 1993, and an Award of Excellence from Bellcore for his work on the digital subscriber line in 1987. He was the Editor-in-Chief of the IEEE TRANSACTIONS ON MULTIMEDIA and a Distinguished Lecturer of the IEEE Circuits and Systems Society from 2000 to 2001. He was the General Co-chair of the Visual Communications and Image Processing 2000 Conference. He was the Editor-in-Chief of the IEEE TCSVT from 1995 to 1997. From 1988 to 1991, he was the chairman of the IEEE CAS Standards Committee and established the IEEE Inverse Discrete Cosine Transform Standard.