Guest Editorial Introduction to the Special Issue on Group and Crowd Behavior Analysis for Intelligent Multicamera Video Surveillance

DESPITE significant progress in human behavior analysis over the past few years, most of today's state-of-the-art algorithms focus on analyzing individual behavior in a simple environment monitored by a single camera. Recently, the widespread availability of cameras and a growing need for public safety have shifted the attention of researchers in video surveillance from individual behavior analysis to group and crowd behavior analysis in multicamera networks. Group behavior analysis provides a novel level for describing events, which are semantically more meaningful, highlighting barely visible relational connections among people. Crowd behavior analysis can also be used for anomaly detection such as panic scenarios, dangerous situations, and illegal behaviors in public spaces.

Although the studies on group and crowd behavior analysis in multicamera networks are valuable for both research and industry, many fundamental problems are unsolved so far, such as how to calibrate multiple cameras, how to compute the topology of camera networks, how to detect moving objects in dynamic scenes, how to track human across camera views, how to complement RGB and depth data for accurate human pose estimation, how to fuse information from multiple cameras for the analysis of group and crowd behavior, how to automatically learn abnormal behavior patterns, just to mention a few of a diverse range of challenges.

The 24 papers in this Special Issue provide comprehensive coverage of the current advances in the area. We hope they can be useful as insights into new directions in group and crowd behavior analysis from videos. We also expect that this Special Issue will motivate more researchers to contribute to this important research area. The papers presented in this special issue are grouped into five main clusters: 1) person localization and reidentification; 2) person tracking; 3) crowd behavior recognition and detection; 4) action clustering and recognition; and 5) abnormal behavior patterns learning and detection. In the following, we present the five clusters in detail.

A. Person Localization and Tracking

In a multicamera surveillance environment, acquiring the trajectories of moving people by localizing and tracking people is very important for subsequent high-level recognition and understanding tasks. In the paper titled "Accelerating Vanishing Point-Based Line Sampling Scheme for Real-Time People Localization," Liu *et al.* propose a fast multicamera people localization approach that is capable of locating a crowd of dense people and estimating their heights in a fairly short time with high accuracy. In the second paper, titled "Robust Visual Tracking via Basis Matching," Zhang *et al.* propose an object tracking method based on basis matching. In the third paper, titled "Globality-Locality-Based Consistent Discriminant Feature Ensemble for Multicamera Tracking," Nithin and Brémond propose an online multicamera multitarget tracking framework that performs adaptive tracklet correspondence by analyzing and understanding contents and properties of video. In the paper titled "Tracking Social Groups Within and Across Cameras," Solera *et al.* propose a method for tracking groups from single and multiple cameras with disjointed fields of view. They formulate the problem as a supervised clustering problem in which a structural SVM classifier learns a similarity measure that is appropriate for group entities. In the last paper, titled "Online Scheme for Multiple Camera Multiple Target Tracking Based on Multiple Hypothesis Tracking," Yoo *et al.* propose an online tracking algorithm for multiple target tracking with multiple cameras.

B. Person Reidentification

Person reidentification aims to match a specific individual across nonoverlapping cameras, which is an important but challenging task in multicamera video surveillance. In this issue, there are six works on person reidentification. In the first paper, titled "An Ensemble of Invariant Features for Person Reidentification," Lee *et al.* propose an ensemble of invariant features (EIFs), which can properly handle the variations of color difference and human poses/viewpoints for matching pedestrian images observed in different cameras with nonoverlapping field of views. In the second paper, titled "Person Reidentification Using Multiple Egocentric Views," Chakraborty *et al.* present a person reidentification framework designed for a network of multiple wearable devices, which builds on commonly used facial feature extraction and similarity computation methods between camera pairs and utilizes a data association method to yield globally optimal and consistent reidentification results with much improved accuracy. In the third paper, titled "PRISM: Person Reidentification via Structured Matching," Zhang *et al.* propose a person reidenficiation method via structured matching (PRISM), an SM method to jointly account for the existing challenges. In the fourth paper, titled "DeepList: Learning Deep Features With Adaptive Listwise Constraint for Person Reidentification," Wang *et al.* propose to learn deep representations with

Digital Object Identifier 10.1109/TCSVT.2017.2669658

an adaptive margin listwise loss for person reidentification. In the fifth paper, titled "Person Reidentification Using Deep Convnets With Multitask Learning," McLaughlin *et al.* train a deep convolutional network to represent a person's appearance as a low-dimensional feature vector that is invariant to common appearance variations encountered in the reidentification problem. In the last paper, titled "From the Lab to the Real World: Re-identification in an Airport Camera Network," Camps *et al.* design and deploy a human reidentification system in a demanding real-world environment: a busy airport.

C. Action Clustering and Recognition

Action recognition is one of the most challenging tasks in video surveillance because of the diversity of video scenes and actions. In the first paper, titled "Laplacian LRR on Product Grassmann Manifolds for Human Activity Clustering in Multicamera Video Surveillance," Wang *et al.* propose a novel representation for multicamera video data to model video sequences as points on the Grassmann manifold and integrate them as a whole in the product manifold form. In the second paper, titled "Two-Stream Dictionary Learning Architecture for Action Recognition," Xu *et al.* propose a novel method based on the two-stream dictionary learning architecture for human action recognition. The architecture consists of interest patch (IP) detector and descriptor, two-stream dictionary models, and a support vector machine (SVM) for classification. In the third paper, titled "Kernel Regularized Data Uncertainty for Action Recognition," Feng and Zhou propose four classifiers for action recognition. To improve the classification performance, the proposed classifiers not only generate the virtual sample but also utilize the Tikhonov matrix to encode the importance of each sample.

D. Crowd Behavior Recognition and Detection

Crowd behavior detection and recognition play very important roles in real-word video surveillance. In recent years, there has been much significant progress achieved by the community. This issue has seven papers on this topic. In the first paper, titled "Crowd Behavior Analysis Using Local Mid-Level Visual Descriptors," Fradi *et al.* propose a compact local representation of the crowd from which a rich set of visual descriptors are extracted for crowd behavior analysis. In the second paper, titled "Detection of Global and Local Motion Changes in Human Crowds," de Almeida *et al*. propose a computer vision method to identify motion pattern changes in human crowds that can be related to an unusual event. In the third paper, titled "Crowded Scene Understanding by Deeply Learned Volumetric Slices," Shao *et al*. build a large-scale crowd data set with 10 000 videos from 8257 scenes, and propose 94 crowd-related attributes for crowded scene understanding. In the fourth paper, titled "Evaluation of Low-Level Features for Real-World Surveillance Event Detection," Xian *et al.* systematically evaluate seven different types of low-level spatiotemporal features in the context of surveillance event detection (SED) using a uniform experimental setup. In the fifth paper, titled "Exploring Coherent Motion Patterns via Structured Trajectory Learning for Crowd Mood Modeling," Zhang *et al.* propose a novel crowd representation termed crowd mood. Crowd mood is established based upon the discovery that the social emotional hypothesis of crowd behaviors can be revealed by investigating the spacing interactions and the structural levels of motion patterns in crowds. In the sixth paper, titled "Online Human Interaction Detection and Recognition with Multiple Cameras," Motiian *et al.* introduce a modeling framework for the online detection and recognition of human interactions from multiple cameras. In the seventh paper, titled "Opportunistic Image Acquisition of Individual and Group Activities in a Distributed Camera Network," Ding *et al.* describe an active camera parameter selection method for capturing high resolution images of targets, when certain events occur, while maintaining coverage of a large area using a camera network.

E. Abnormal Behavior Patterns Learning and Detection

Abnormal behavior patterns detection from video is a hot topic for many video surveillance applications. In the first paper, titled "Histograms of Optical Flow Orientation and Magnitude and Entropy to Detect Anomalous Events in Videos," Colque *et al.* propose an approach for detecting anomalous events in videos with crowds, which uses general concepts such as orientation, velocity, and entropy to capture anomalies. In the second paper, titled "Toward Abnormal Trajectory and Event Detection in Video Surveillance," Coşar *et al.* propose a unified approach for abnormal behavior detection and group behavior analysis in video scenes, which incorporates the output of object trajectory analysis and pixel-based analysis for abnormal behavior inference. In the last paper, titled "Discriminative Dictionary Learning With Motion Weber Local Descriptor for Violence Detection," Zhang *et al.* propose a more effective and efficient algorithm for detecting violence from motion images based on an improved Weber local descriptor.

The success of this Special Issue would not have been possible without the contribution and support from many individuals. We thank all the authors who submitted papers to this Special Issue. We thank the reviewers for their effort and time in reviewing each manuscript in a timely and professional way. We thank all the administrative staff for their enormous and valuable assistance. Last, but not least, we would like to extend thanks to the Editor-in-Chief, Prof. Dan Schonfeld, for his valuable advice and strong support in making this Special Issue possible.

HONGXUN YAO
Harbin Institute of Technology
Harbin, China

ANDREA CAVALLARO
Queen Mary University of London
London, U.K.

THIERRY BOUWMANS
University of La Rochelle
La Rochelle, France

ZHENGYOU ZHANG
Microsoft Research
Redmond, WA, USA

Hongxun Yao (M'11) received the B.S. and M.S. degrees in computer science from Harbin Shipbuilding Engineering Institute, Harbin, China, in 1987 and 1990, respectively, and the Ph.D. degree in computer science from Harbin Institute of Technology, Harbin, China, in 2003. She is a Professor with the School of Computer Science and Technology, Harbin Institute of Technology. She has authored or co-authored six books and over 200 scientific papers. She received the honor title of New Century Excellent Talent in China. Her research interests include computer vision, pattern recognition, multimedia computing, and human–computer interaction technology.

Andrea Cavallaro received the Ph.D. degree in electrical engineering from Swiss Federal Institute of Technology, Lausanne, in 2002. He is a Professor of Multimedia Signal Processing and the Director of the Centre for Intelligent Sensing, Queen Mary University of London, U.K. He has published over 150 journal and conference papers, one monograph on video tracking (Wiley, 2011), and three edited books: *Multi-Camera Networks* (Elsevier, 2009); *Analysis, Retrieval, and Delivery of Multimedia Content* (Springer, 2012); and *Intelligent Multimedia Surveillance* (Springer, 2013). He was a Research Fellow with British Telecommunications from 2004 to 2005. He was a recipient of the Royal Academy of Engineering Teaching Prize in 2007; three student paper awards on target tracking and perceptually sensitive coding at the IEEE ICASSP in 2005, 2007, and 2009; and the best paper award at IEEE AVSS 2009. He is a Senior Area Editor of IEEE TRANSACTIONS ON IMAGE PROCESSING and an Associate Editor of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY and *IEEE Multimedia*.

Thierry Bouwmans is an Associate Professor with University of La Rochelle, La Rochelle, France. His research interests include the detection of moving objects in challenging environments as it is a pre-step for behavior detection in video surveillance. He has supervised five Ph.D. students in his research fields. He has authored over 50 papers in refereed international journals and conferences and has co-edited two books in background/foreground separation for video surveillance, and robust PCA via decomposition into low rank and sparse matrices. His research particularly investigated the use of fuzzy concepts, discriminative subspace learning models, and robust PCA in video surveillance. He has been the Lead Guest Editor of the Special Issue on Background Modeling for Foreground Detection in Real-World Dynamic Scenes in the journal *Machine Vision and Applications*. He is the Co-ordinator of the BGS Library and LRS Library. He is the Administrator of the Background Subtraction Web Site (33,115 visits and 17,636 visitors). He is also a Reviewer for prestigious international journals, including IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON MULTIMEDIA, and IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, and journals for Springer and Elsevier, and top-level conferences such as CVPR, ICPR, ICIP, and AVSS.

Zhengyou Zhang (F'05) received the B.S. degree in electronic engineering from Zhejiang University, Hangzhou, China, in 1985; the M.S. degree in computer science from University of Nancy, Nancy, France, in 1987; and the Ph.D. degree in computer science and the D.Sc. (*Habilitation à Diriger des Recherches*) degree from University of Paris XI, Paris, France, in 1990 and 1994, respectively. He is a Principal Researcher and Research Manager with Microsoft Research, Redmond, WA, USA. He was a Senior Research Scientist with INRIA French National Institute for Research in Computer Science and Control, France, and spent a one-year sabbatical as an Invited Researcher with the Advanced Telecommunications Research Institute International, Kyoto, Japan. He has published over 250 papers in refereed international journals and conferences, and has co-authored five books. He is an Inventor of over 130 issued patents. He is an ACM Fellow. He received the IEEE Helmholtz Test of Time Award at ICCV 2013 for his paper published in 1999 on camera calibration, currently known as Zhang's method. He has received eight Ship-It Awards at Microsoft. He is also the Founding Editor-in-Chief of IEEE TRANSACTIONS ON AUTONOMOUS MENTAL DEVELOPMENT and an Area Editor of *Journal of Computer Science and Technology*. He served as an Associate Editor of IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE TRANSACTIONS ON MULTIMEDIA, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the *International Journal of Computer Vision*, the *Machine Vision and Applications* journal, and *International Journal of Pattern Recognition and Artificial Intelligence*. He is on the Honorary Board of *International Journal of Computer Vision* and on the Steering Committee of *Machine Vision and Applications*. He is serving as a General Chair at the IEEE Conference on Computer Vision and Pattern Recognition 2017.