

“Signal, image and video processing” special issue: Semantic representations for social behavior analysis in video surveillance systems

Shengping Zhang · Huiyu Zhou · Baochang Zhang ·
Zhenjun Han · Yuliang Guo

Published online: 13 November 2014
© Springer-Verlag London 2014

Video surveillance has been attracting increasing attentions in the computer vision community because of its wide industrial applications and important scientific values. Among the related topics, automatic behavior analysis plays an extremely important role and has witnessed tremendous progress in the last twenty years. Recently, researchers in video surveillance shift their attention from the monitoring of a single person’s behaviors in a relatively simple environment to that of social behavior of multiple persons in crowded environments. In contrast to single person’s behavior, social behavior analysis faces more challenges such as complex interaction, diverse semantics and various expressions. This is due to the gap between the information directly extracted from videos and semantic interpretations by our human beings.

To bridge this gap, a number of feature representation approaches (e.g., Cuboids, HOG/HOF, HOG3D and eSURF) have been subsequently reported to address the coherence between the extracted features and the semantic interpretations. Unfortunately, due to the redundancy and complexity,

these hard-crafted features may lead to diverse variations of semantic representations for social behavior analysis.

In recent years, novel semantic representations have proven to be an effective tool for social behavior analysis. For example, social force model and its variant have proven to perform well in social behavior recognition. Such high-level semantic representations achieve desired performance even if in crowded environments. Besides, statistical approaches, syntactic approaches and description-based approaches also gain increasing attention in computer vision community.

The primary purpose of this special issue is to organize a collection of recently developed high-level semantic representations for social behavior analysis, spreading over motion trajectory acquisition and analysis, semantic feature extraction, social behavior analysis and applications. It provides an international forum for researchers to report and share the recent original developments in this field in an original research paper style. This special issue includes 13 papers which have been accepted in the categorization of three themes: visual object tracking, pedestrian detection and recognition, and behavior analysis and recognition. See details as follows:

- (1) Visual object tracking: This is an active research topic in the last decade due to its wide applications in intelligent video surveillance system. The major challenge in the recent development is how to maintain the performance of the systems in partial occlusions, pose or illumination changes. In this special issue, Gai and Luo proposed to extract geometric features based on reduced quaternion wavelet transform by using color space information. Sun, Zhang, Xie, Gao, Wang and Heidingsfelder propose an active mating-based visual tracking by exploring the biologically inspired color surface coding to refine the original interest point detector, which benefits for robust rep-

S. Zhang (✉)
Harbin Institute of Technology, Weihai, China
e-mail: s.zhang@hit.edu.cn

H. Zhou
Queen’s University Belfast, Belfast, UK
e-mail: h.zhou@ecit.qub.ac.uk

B. Zhang
Beihang University, Beijing, China
e-mail: bczhang@buaa.edu.cn

Z. Han
University of Chinese Academy of Sciences, Beijing, China
e-mail: hanzhj@ucas.ac.cn

Y. Guo
Brown University, Providence, RI, USA
e-mail: yuliang_guo@brown.edu

resentation to extract the suitable interior object areas for object matting. Sun, Yao and Lu propose to fuse multiple cues in a particle filter framework for visual tracking, which helps to seize the most informative properties of distinguishing the target from background. Chen, Zhang, Ruan, Xu, Sun, Gong, Min and Lei propose a tracking method based on mean shift by exploring simultaneously the temporal and spatial information of the tracked object. A cascade classification method based on nearest neighbor and SOM (Self-Organizing Map) is employed as a confirmation step to eliminate spurious objects. The forward and backward tracking results are further combined to improve the localization accuracy and tolerate at the same time scale variation. Madrigal, Hayet and Lerasle present an interacting multiple pedestrian tracking method for monocular systems that incorporates a prior knowledge about the environment and about interactions between targets.

- (2) **Pedestrian detection and recognition:** Detecting the pedestrian and then recognizing it is useful to identify the specific pedestrian in a scene over multiple cameras without overlapped views. Due to low contrast among different pedestrians, pedestrian recognition is very challenging. The accepted papers in this special issue make some progress to overcome those challenges. Lin, Zeng and Huang propose a method to combine a cell-structured HOG feature and adaptive local binary pattern feature to solve the problem that HOG is vulnerable to the interference of vertical background gradient information in pedestrian detection. Yang, Delp and Du propose a categorization-based two-stage pedestrian detection system to efficiently locate pedestrians within our collected TASI 110-car naturalistic driving dataset. Category information including vehicle status, location and time is automatically extracted, and efficient category-specific detection algorithms are designed for different scenario categories. Qin, Deng and Yung propose a scene categorization method based on local-global feature fusion and multi-scale multi-spatial resolution encoding with the bag-of-contextual-visual-word (BOCVW) models. Yu, Gan, Yang, Ding, Jiang, Wang and Li improve the traditional local binary patterns by utilizing the least square estimate technique to optimize the weight and minimize the local absolute difference, which leads to more stable directional features. In addition, a novel rotation invariant texture classification approach is presented.

- (3) **Behavior analysis and recognition:** The final purposes of a vision system are to analyze and recognize the behaviors of the objects in a scene. The behaviors of interest include both individual and crowded behaviors. Aiming at the effective, accurate and freely used hand gesture recognition with Kinect, Jiang, Wu, Gao, Zhao and Kung present a viewpoint-independent hand gesture recognition system. Firstly, based on the rules about gesturer's posture under optimal viewpoint, the gesturer's point clouds are built and transformed to the optimal viewpoint with the exploration of the joint information. Then, Laplacian-based contraction is used to extract skeletons from point clouds of hands. A novel partition-based algorithm is further proposed to recognize the gestures. Zhang, Huang, Qin, Zhao, Yao and Xu address the problem of dense crowd event recognition in the surveillance video and present a novel crowd behavior representation called Bag of Trajectory Graphs (BoTG). Chen, Ye, Zou, Li, Cui and Jiao propose to represent trajectories with "bag-of-word" features from a spatially distributed codebook, and then use a Replicated Softmax model to characterize trajectories with latent topic units. By stacking an additional label layer or representation layers, the Replicated Softmax model is enhanced with discrimination and generalization capability. Threat detection is a challenging problem, because threats appear in many variations and differences to normal behavior and can be very subtle. Burghouts, Schutte, Hove, Broek, Baan, Rojas, Huis, Rest, Hanckmann, Bouma and Sanroma consider threats on a parking lot, where theft of a truck's cargo occurs. The novelty of this paper is an encoding of these threat observables in a semantic, intermediate-level representation, based on low-level visual features that have no intrinsic semantic meaning themselves. The semantic representation encodes the notions of trajectories, zones and activities.

Acknowledgments The Guest Editors would like to thank all the authors for submitting their high-quality manuscripts to this special issue and all the reviewers for providing their timely review comments and suggestions. They would also like to thank the Editor-in-Chief Prof. A. Enis Cetin and the journal managers Neeraja Prabakaran for encouraging and supporting this special issue.