

## Editorial

# Video Analysis for Human Behavior Understanding

Jenq-Neng Hwang,<sup>1</sup> Changick Kim,<sup>2</sup> and Hsu-Yung Cheng<sup>3</sup>

<sup>1</sup>University of Washington, Seattle, WA 98195-5850, USA

<sup>2</sup>Korea Advanced Institute of Science and Technology, Daejeon 305-701, Republic of Korea

<sup>3</sup>National Central University, Taoyuan County 32001, Taiwan

Correspondence should be addressed to Jenq-Neng Hwang, hwang@u.washington.edu

Received 10 June 2010; Accepted 10 June 2010

Copyright © 2010 Jenq-Neng Hwang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Video cameras are becoming increasingly ubiquitous and pervasive in our daily life. Along with the fast growing number of exchanged and archived videos, there is an urgent need for advanced video analysis techniques that can systematically interpret and understand the semantics of video contents, within the application domains of security surveillance, intelligent transportation, health/home care, video indexing and retrieving, video summarization and highlighting, and so on. Understanding human behaviors based on video analysis calls for even greater challenges due to very large variations of human bodies and their motion activities under all kinds of contexts such as different viewing perspectives, dynamically changing backgrounds, dressing colors, changing human poses, human-human occlusions, and body parts self-occlusions. To overcome these challenges, not only the traditional image processing, computer vision, pattern recognition, and machine learning techniques are required, but also advanced estimation theory and statistical inference, articulated 2D/3D human body modeling and synthesis, sophisticated database or rules for events/behaviors, and so on are critically desired.

The primary focus of this special issue is on the advanced video analysis and machine learning techniques for understanding human behaviors, starting from human object detection, segmentation and tracking, 2D/3D spatial and temporal features extraction, 2D/3D human body modeling and synthesis, event discovery and behavior learning, extension to crowd behavior analysis, system performance evaluation, and potential applications of these techniques. We have received an impressive number of paper submissions for this special issue, and thanks to the tremendous efforts

of many responsible reviewers, we were able to select twelve very high quality papers, which are organized and ordered according to three broadly categorized topics: (1) human object detection, segmentation, and tracking; (2) human body modeling and action recognition; (3) crowd estimation and crowd behavior analysis. In the following, we will briefly summarize these accepted papers.

Papers 1–4 mainly address the issues of *human object detection, segmentation and tracking*. The paper by Conte et al., entitled “*An experimental evaluation of foreground detection algorithm in real scenes*,” conducts a thorough experimental comparison of four notable foreground detection algorithms, using quantitative performance indices on a large dataset of videos covering several realistic applicative scenarios. This paper concludes that both *mixture of Gaussians* (MOG) and *enhanced background subtraction* (EBS) algorithms are quite versatile and can be used effectively in most situations. The *self-organizing background subtraction* (SOBS) algorithm gives good results in indoor environments but can have some problems in outdoor settings. Finally, *statistical background algorithm* (SBA) is consistently inferior to the others, so its adoption is not advisable. The paper by Ahn et al., entitled “*Automatic moving object segmentation from video sequences using alternate flashing system*,” jointly considers sensitivity, color, coherence, and smoothness for segmentation. The proposed algorithm employs a flashing system to obtain a series of lit and unlit frames from a single camera. By comparing the unlit and lit frames, a sensitivity map which provides depth cues can be constructed. Moreover, real-time segmentation can be achieved with an acceleration mechanism. The paper by Chang

et al., entitled “*Localized detection of abandoned luggage*,” explores the automatic detection of abandoned luggage by using *foreground-mask sampling* to detect luggage with arbitrary appearance and *selective tracking* to track owners. More specifically, the object of interest is localized through foreground-mask sampling, which enables tracking to be performed in a more selective and localized manner. The detected foreground region is checked to determine whether it is a human via a combination of skin color information and body contours. If the region is identified as a human, the region is discarded. If not, it is assumed to be a luggage item. A local search region is constructed around the detected luggage to see whether its owner is in close proximity in the current frame. If the owner is found, the region is again discarded because the owner exhibits no intention of abandoning the luggage. The paper by Wu et al., entitled “*A hierarchical estimator for object tracking*,” integrates local and global estimates into one joint estimate for mutual compensation. A feedback loop is implemented to achieve iterative optimization and gain improvement from both the local and global mutual compensation. Such integration allows the tracker to adjust fusion gain optimally based on the environment conditions and therefore to obtain higher tracking accuracy.

Papers 5–10 research the issues of *human body modeling and action recognition*. In “*Novel kernel based recognizers of human actions*,” Danafar et al. study unsupervised and supervised recognition of human actions in video sequences. The videos are represented by probability distributions and then meaningfully compared in a probabilistic framework. They introduce two novel approaches, which outperform the state-of-the-art algorithms when tested on the KTH and Weizmann public datasets: one is an unsupervised non-parametric kernel-based method that exploits the Maximum Mean Discrepancy test statistic and the other is a supervised method based on a support vector machine (SVM) with a characteristic kernel specifically tailored to histogram-based information. The paper by Yu et al., entitled “*Efficient Human Action and Gait Analysis Using Multiresolution Motion Energy Histogram*,” effectively adopts the average motion energy (AME) image to describe human motions by proposing a histogram based approach to improve the computation efficiency. In the paper, the human action/gait recognition problem is formulated as a histogram matching problem, and a quadtree decomposition-based multiresolution structure on the motion energy histogram (MEH) is used to speed up the calculation. Two applications, action recognition and gait classification, are conducted in the experiments to demonstrate the feasibility and validity of the proposed approach. In the paper, “*Recognizing human actions using NWFE-based histogram vectors*,” Lin et al. presented a novel system for human action recognition. Two research issues are addressed in this paper, namely, *motion representation* and *subspace learning*. The proposed system first extracts a combined feature, which integrates the signal distance feature and the width feature extracted from a human pose silhouette. Principle component analysis (PCA) is then employed to reduce the dimensionality of feature vectors and the k-means algorithm is also applied to

construct a codebook. Finally, a Bayesian classifier is utilized to label NWFE-based histogram vectors.

This paper by Kim et al., entitled “*Human action recognition using ordinal measure of accumulated motion*,” presents a robust method for recognizing human actions from a single query action video with variations of appearances. An accumulated motion image (AMI) is introduced and resized based on intensity averaging so as to create a rank matrix. The human action can be recognized through computing the distances from the rank matrix of the query action video to the rank matrices of all local windows in the reference videos. The proposed method does not require any preprocessing task such as learning and segmentation, and therefore it is quite efficient and robust. The paper by Wang et al., entitled “*A two-stage bayesian network method for 3D human pose estimation from monocular image sequences*,” designs a 3D pose estimator that uses a two-stage inference hierarchy. A simulated annealed mechanism is utilized to infer the maximum posteriori distributions of joint positions. The two-stage approach demonstrates improvement on 3D pose estimation accuracy. Moreover, both 2D and 3D pose estimations can be obtained by this two-stage framework effectively. In addition to body motion analyses for human behavior understanding, facial expression can also provide a useful indicator for behaviors. The paper by Lee et al., entitled “*Facial affect recognition using regularized discriminant analysis-based algorithms*,” proposes a regularized discriminant analysis-based boosting algorithm (RDAB) with effective Gabor features to recognize the facial expressions, such as happiness, disgust, fear, anger, sadness, surprise, and neutral state. The RDAB combines strengths of linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA), and the optimal parameters used in the RDAB are estimated via the particle swarm optimization (PSO) algorithm.

When dealing with visual based human behavior analyses, it is inevitable to take into account the behaviors of a group of people in the same scene. This leads us to the *crowd estimation and crowd behavior analysis*, which is becoming more and more attractive as we see the rapid deployment of wide-area surveillance cameras. Two papers (Papers 11-12) are selected for this category. The paper by Conte et al., “*A method for counting moving people in video surveillance videos*,” presents a method based on the use of SURF features and of an epsilon support vector regressor to provide an estimate of people count. The algorithm also takes into account problems due to partial occlusions and viewing perspective. The proposed method is favorably compared with the state-of-the-art algorithm with an improved accuracy while retaining the robustness. In the article “*Robust Recognition of Specific Human Behaviors in Crowded Surveillance Video Sequences*,” Takahashi et al. describes a method that can detect specific human behaviors even in crowded surveillance video scenes. The system recognizes specific behaviors based on the trajectories created by detecting and tracking people in a video. It detects people using a histogram of gradient (HOG) descriptor and an SVM classifier, and it tracks the regions by calculating the 2D color histograms. The implemented system precisely

identifies specific behaviors and achieves the first place for detecting running people in the TRECVID 2009 Surveillance Event Detection Task.

We hope that this special issue provides readers with a better understanding of automated human behavior analysis, from the visual signal processing and machine learning perspective. We also hope these accepted papers can provide avenues and inspirations for further exploration within the research community. In conclusion, we would like to show our deep appreciations to all the submitting authors for their terrific efforts in preparing the manuscripts for this special issue. We also thank so many reviewers for their valuable feedback throughout the review process and tremendous contribution toward producing the highest quality of this special issue. Finally, we would like to thank Professor Phillip Regalia, the Editor-in-Chief, and the editorial board members of the EURASIP Journal of Advances in Signal Processing, for their approval and comments on the development of this proposed special issue.

*Jenq-Neng Hwang  
Changick Kim  
Hsu-Yung Cheng*