

Editorial

Human-Activity Analysis in Multimedia Data

A. Enis Cetin,¹ Eric Pauwels,² and Ovidio Salvetti³

¹ *Department of Electrical and Electronics Engineering, Bilkent University, Ankara 06800, Turkey*

² *Signals and Images Research Group, Centre for Mathematics and Computer Science (CWI), 1098 SJ Amsterdam, The Netherlands*

³ *Institute of Information Science and Technologies (ISTI), Italian National Research Council (CNR), 56124 Pisa, Italy*

Correspondence should be addressed to A. Enis Cetin, cetin@bilkent.edu.tr

Received 11 November 2007; Accepted 11 November 2007

Copyright © 2008 A. Enis Cetin 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.

Many important applications in multimedia revolve around the detection of humans and the interpretation of their behavior. These include surveillance and intrusion detection, video conferencing applications, assisted living applications, and automatic analysis of sports videos, broadcasts, and movies, to name just a few. Success in these tasks often requires the integration of various sensor or data modalities such as video, audio, motion, and accompanying text, and typically hinges on a host of machine-learning methodologies to handle the inherent variability and complexity of the ensuing features. The computational efficiency of the resulting algorithms is critical since the amount of data to be processed in multimedia applications is typically large, and in real-time systems, speed is of the essence.

There have been several recent special issues dealing with the detection of humans and the analysis of their activity relying solely on video footage. In this special issue, we have tried to provide a platform to contributions that make use of a broader spectrum of multimedia information, complementing video with audio or text information as well as other types of sensor signals, whenever available.

The first group of papers in the special issue addresses the joint use of audio and video data. The paper “Audio-visual head orientation estimation with particle filtering in multisensor scenarios” by C. Canton-Ferrer et al. describes a multimodal approach to head pose estimation of individuals in environments equipped with multiple cameras and microphones such as smart rooms for automatic video conferencing. The fusion of audio and vision is based on particle filtering.

S. Shivappa et al., in the paper “An iterative decoding algorithm for fusion of multimodal information,” present an algorithm for speech segmentation in a meeting room scenario using both audio and visual cues. The authors put for-

ward an iterative fusing algorithm that takes advantage of the theory of turbo codes in communications theory by establishing an analogy between the redundant parity bits of the constituent codes of a turbo code and the information from different sensors in a multimodal system. Dimoulas et al., in the paper “Joint wavelet video denoising and motion activity detection in multimodal human activity analysis: application to video-assisted bioacoustic/psychophysiological monitoring,” also integrate both audio and video information to develop a video-assisted biomedical monitoring system that has been tested for the noninvasive diagnosis of gastrointestinal motility dysfunctions.

The articles by N. Ince et al., titled “Detection of early morning daily activities with static home and wearable wireless sensors,” and B. Toreyin et al., titled “Falling person detection using multisensor signal processing,” are concerned with indoor monitoring and surveillance applications that rely on the integration of sensor data. N. Ince et al. describe a human activity monitoring system to assist patients with cognitive impairments caused by traumatic brain injury. The article details how fixed motion sensors combined with accelerometer embedded in wearable wireless sensors allow the system to detect and classify daily morning activity. B. Toreyin et al. outline a smart room application employing passive infrared and vibration sensors, as well as audio, to reliably detect a person falling.

The rest of the papers in this issue describe video-based surveillance applications. F. Porikli et al., in the paper “Robust abandoned object detection using dual foregrounds,” detect abandoned objects by estimating dual foreground images from video recorded in an intelligent building. G. Pieri and D. Moroni, in the paper “Active video-surveillance based on stereo and infrared imaging,” describe a video surveillance system integrating information from regular stereo and

infrared cameras. They exploit the strengths of both modalities by utilizing the more accurate localization made possible by the stereo cameras in combination with the improved detection robustness that results from inspecting the IR data. The article by L. Raskin et al., titled “Using gaussian process annealing particle filter for 3D human tracking,” tracks humans in 3D scenes using particle filters. The article by M. Hossain et al., titled “Edge segment-based automatic video surveillance,” describes a paper using image edge information for automatic video surveillance.

*A. Enis Cetin
Eric Pauwels
Ovidio Salvetti*