Special issue on multimedia Surveillance systems: guest editorial

J. K. Aggarwal · Rita Cucchiara

1 Introduction

It is with considerable pride that we present this special issue of ACM multimedia based on the presentations at the third Video Surveillance and Sensor Network workshop, in conjunction with the ACM conference in Singapore 2005. The papers were thoroughly reviewed independently of the review process for the workshop. This special issue consists of eight papers drawn from a number of areas. It appears that we are breaking new ground as explained in this issue.

Whenever we say multimedia, we think of systems and services that manage heterogeneous data for human-oriented applications; human users are normally the subjects who access and use multimedia data, multimedia streams, multimedia content, and multimedia interfaces in many different applications contexts.

Following this abstraction, *multimedia surveillance systems* would be only a surveillance system able to produce output of the task in a multimedia format, providing distilled video, images and sounds of the monitored environment, which would possibly be annotated in an efficient and standard way or possibly transcoded in another media such as text or animation, to improve further querying to surveillance stored data.

Instead, the idea of multimedia surveillance systems must be enlarged. It is not only a system capable of furnishing multimedia data, but also collecting, processing in real-time, correlating and handling multimedia data coming from different sources. Video surveillance systems should include multimodal video sources: distributed fixed cameras, PTZ, moving and omni-directional cameras, and multi-spectral cameras (e.g., thermal cameras). Multimedia surveillance systems should improve visual data with audio streams and information coming from other sensors. In large distributed environments, the exploitation of networks of small cooperative sensors should dramatically improve the surveillance capability to a few higher levels.

This special issue addresses works in the field of the new generation of computer-based multimedia surveillance systems, with surveillance modules and sensors networks co-operating in order to create future processing and understanding systems of surveillance data.

2 Scanning the issue

The deployment of multiple sensors for surveillance applications is a key research topic. Having numerous sources of information to be processed poses several problems.

First of all, the deployment of multiple cameras in a scene requires finding the optimal positions and poses for each camera, to achieve the best trade-off between number of cameras (i.e., costs) and coverage of the scene.

In the paper entitled "Calibrating and Optimizing Poses of Visual Sensors in Distributed Platforms" Hoerster and Lienhart propose an approach suitable for set-ups in which synchronization is not required, cameras can have disjointed fields of view (but with a reasonable

J. K. Aggarwal (✉)
Department of Electrical and Computer Engineering,
University of Texas, Austin, USA
e-mail: aggarwaljk@mail.utexas.edu

R. Cucchiara
Department of Computer Engineering,
University of Modena and Reggio Emilia, Madona, Italy
e-mail: rita.cucchiara@unimo.it

overlap between camera subgroups), and no upper limit on the number of cameras and displays under calibration is imposed. A linear programming approach is used to determine jointly the pan and tilt angle for each camera that maximizes the coverage of the space at a given sampling frequency.

When the camera/sensor optimal deployment has been determined, the calibration of the complete system is required in order to correlate data provided by the different sensors. With this regard, Wren et al. in the paper "Functional Calibration for Pan-Tilt-Zoom Cameras in Hybrid Sensor Networks" propose an efficient method to automatically calibrate a single PTZ camera with a network of one-bit sensors. The target problem is that of a system without complete coverage of the scene, in which gaps between sensor coverage can bring significant problems and loss of information. They propose a hybrid perceptual system based on a model of activity able to recover not the full calibration of the devices, but a functional calibration consisting of a blending of geometry estimation and simple behavioral model discovery.

On a similar topic, the work "Information Assimilation Framework for Event Detection in Multimedia Surveillance Systems" by Atrey et al. describes how to handle a system not composed of synchronized and heterogeneous sensors. In these cases, information assimilation (defined as "how to combine these multiple sources of data") becomes a crucial task. In this paper, a new framework based on a hierarchical probabilistic assimilation approach is depicted. The framework is used to detect atomic and compound events.

From the perspective of the human operator devoted to monitoring a wide and complex scene, having more cameras allows her/him, on one hand, to have complete coverage of the scene, but imposes, on the other hand, a continued and tedious monitoring of many video displays, limiting the reliability and accuracy of such a control. For this reason, Wang et al., in the paper entitled "GPU-Friendly Warped Display for Scope-Maintained Video Surveillance", propose a method for helping the surveillance in monitoring specific areas and the overview of the scene at the same time. The method maps the input video as a texture on a deformed mesh, depending on the position and shape of the region of interest. Mesh deformation and texture mapping are achieved in real time by exploiting the speed up available in modern graphics processing units (GPUs).

Besides the problem of having effective monitoring of the scene by means of human operators, distributed video surveillance also calls for automatic techniques able to detect, track, and analyze moving objects (in particular, pedestrians) in the scene. The work

"Constructing Task Visibility Intervals for a Surveillance System" by Ser-Nam Lim et al. addresses, for instance, the problem of constructing the "task visibility intervals" in a system composed of multiple PTZ cameras. These intervals contain information about what can be sensed in the future time intervals. Such a construction requires a prediction regarding future object motion taking into consideration object occlusions and camera control parameters. The paper entitled "Surveillance Camera Scheduling: A Virtual Vision Approach" by Zubair Qureshi and Terzopoulos, instead, exploits a virtual environment simulator to evaluate an active camera system in which a PTZ camera is guided by the analysis of several static camera inputs in order to acquire high-resolution images of moving pedestrians. Simulated video streams are used to efficiently test the proposed approach.

Finally, a significant effort has been made in recent years to find both new media to be analyzed to infer data about the scene and algorithms for high-level analysis of the scene. The work "Event Detection in an Audio-Based Sensor Network" by McHugh and Smeaton belongs to the first class and proposes to exploit audio data to detect events. Their analysis based on the mean of the volume, the zero-crossing rate, and the frequency tries to correlate these data with events detected visually. The results presented in the paper demonstrate that detecting events based on their volume only returned satisfactory results. The work "Classifying Spatiotemporal Object Trajectories Using Unsupervised Learning in the Coefficient Feature Space" by Naftel and Khalid, instead, belongs to the second class, by proposing a Mahalanobis classifier to cluster and classify trajectories and detect anomalous behaviors.

3 Conclusion

Visual surveillance and sensor networks are in their early stage of evolution. These papers are state of the art at the moment. Certainly, many more challenges lie ahead in order to make the systems viable for prime-time.

J. K Aggarwal received his Ph.D. in 1964 in Electrical Engineering at the University of Illinois, Urbana. He has been on the faculty of the University of Texas at Austin since 1964. He is currently one of the Cullen Professors of Electrical and Computer Engineering at the University of Texas at Austin.

His research interests include image processing, computer vision, and pattern recognition. The current focus of his research is on the automatic recognition of human activity and interactions in video sequences, and on the

use of perceptual grouping for the automatic recognition and retrieval of images and videos from databases.

Dr. Aggarwal is a Life Fellow of IEEE, Golden Core member of the IEEE Computer Society, Fellow of IAPR, and Fellow of AAAS. He is the recipient of the 2004 K.S. Fu Prize of the International Association for Pattern Recognition and the recipient of the 2005 IEEE Kirchmayer Graduate Teaching Award. He is the author or co-author of numerous books, book chapters, papers, proceeding papers, and technical reports.

Rita Cucchiara graduated in Electronic Engineering in 1989 and received the Ph.D. in Computer Engineering in 1991 at University of Bologna, Italy. Currently she is Full Professor at Faculty of Engineering at University of Modena and Reggio Emilia and heads the ImageLab Laboratory in Modena (http://imagelab.ing.unimo.it).

Her current interests are in pattern recognition and computer vision for video surveillance, medical imaging, and multimedia. Video surveillance activity is devoted to object and people segmentation, tracking, and people behavior analysis in indoor and outdoor applications. Rita Cucchiara is responsible for many Italian and International projects. In Medical Imaging, she is working on color and texture analysis for dermatological imaging. In Multimedia she is working in semantic transcoding and video annotation.

Rita Cucchiara is the author of more than 100 papers in journals, and conference proceedings. In 2006, she received the Fellowship of IAPR for contribution in pattern recognition for video-surveillance. She is an IEEE, ACM, and IAPR Member. In 2007, she will be the general chair of the ICIAP2007 Conference in Italy.