Visual Sensor Networks for Infomobility


Abstract—The wide availability of embedded sensor platforms and low-cost cameras—together with the developments in wireless communication—make it now possible the conception of pervasive intelligent systems centered on vision. Such systems may be understood as distributed and collaborative sensor networks, able to produce, aggregate and process images in order to understand the observed scene and communicate the relevant information found about it. In this paper, we investigate the peculiarities of visual sensor networks with respect to standard vision systems and we identify possible strategies to accomplish image processing and analysis tasks over them. Although the rather strong constraints in computational and transmission power of embedded platforms that may prevent the use of state of the art computer vision and pattern recognition methods, we argue that multi-node processing methods may be envisaged to decompose a complex task into a hierarchy of computationally simpler problems to be solved over the nodes of the network. These ideas are illustrated by describing an application of visual sensor network to infomobility. In particular, we consider an experimental setting in which several views of a parking lot are acquired by the sensor nodes in the network. By integrating the various views, the network is capable to provide a description of the scene in terms of the available spaces in the parking lot.

Keywords: image mining, sensor networks, infomobility, object detection, change detection.

DOI: 10.1134/S1054661811010093

1. INTRODUCTION

In the last years, the wide availability of embedded systems and low-cost camera sensors—together with the developments in wireless communication—has made it possible the conception of pervasive intelligent systems centered on image data [1]. Such visual Wireless Sensor Networks (WSNs), employing a great number of low power camera nodes, are a novel vision-based applications thanks to the great informative power of imaging. For example, visual WSNs may be used to monitor in real-time the crowd in shopping mall, airports and stadiums. By mining the scene, the network may detect anomalous and potentially dangerous events [2]. Similarly, visual WSNs may be used for environmental monitoring, for the remote control of elderly patients in e-Health [3] and for human gesture analysis in ambient intelligence applications [4].

In brief, the key feature of visual WSNs is the combination of the versatility and independence from a physical infrastructure typical of general WSNs with the richness of information that can be gained through imaging techniques, computer vision and image understanding. In this sense, a visual WSN may be understood as a distributed and collaborative sensor network, able to produce, aggregate and process images in order to mine the observed scene and communicate the relevant information found about it. In particular, each acquisition node is capable to acquire a view of the scene to be mined. Then, in cooperation with each other, the various devices in the network are able to process and aggregate the acquired views in order to predicate something about the scene.

A successful design and development of such a system cannot be achieved without suitable solutions to the involved computer vision problems. Although the computer vision problems may be still decomposed into basic computational tasks (such us feature extraction, object detection and object recognition), in the context of visual WSNs, it is not directly possible to use all the methods that have been developed to solve such tasks and that are already available in the specific literature [5]. Indeed, since WSNs usually require a large number of sensors, possibly scattered over a large area, the unit cost of each device should be small to make the technology affordable. Therefore, cost constraints limit the computational and transmission power of sensor nodes as well as the fidelity of acquired images,
with consequences on the employable computer vision algorithms. Since untethered sensors are usually supplied by batteries, it is also necessary to analyze carefully the tradeoffs between quality of processing and energy consumption, in order to avoid too frequent battery replacement.

In light of these considerations, performing computer vision tasks over WSNs is somewhat challenging from a technological viewpoint. However, probably, the change in perspective offered by the possibility to perform pervasive computing and by the ductility in organizing the network topology greatly compensates for the current technical limits of visual WSNs. In addition, the topic is intriguing and may open the way to more theoretical investigations. In this paper, we investigate multi-node processing methods that may be envisaged to decompose a complex vision task into a hierarchy of computationally simpler problems, to be solved over the nodes of the network. Besides computational advantages, such hierarchical decision making approach may lead to more robust and fault-tolerant results. We illustrate these ideas by describing an application of visual sensor network to infomobility. To this end, we consider an experimental setting in which several views of a parking lot are acquired by the sensor nodes in the network. By integrating the various views, the network is capable to provide a description of the scene in terms of the available spaces in the parking lot. This paper extends and complements the conference paper [6].

2. BACKGROUND

Embedded vision platform first appeared in connection with video-surveillance and robotics. Thanks to the drop in technology costs, nowadays, their application range has become wider so as to encompass several sectors of public and private life and, in particular, infomobility. We first discuss existing embedded vision platforms and basic features of visual WSNs (Section 2.1); then we survey the specificities of performing image analysis over WSNs, introducing in particular the multi-view vision problem (Section 2.2).

2.1. Visual Sensor Networks

Following the trends in low-power processing, wireless networking and distributed sensing, visual WSNs are experiencing a period of great interest, as shown by the recent scientific production (see e.g. [7]). A visual WSN consists of tiny visual sensor nodes called camera nodes, which integrate the image sensor, the embedded processor and a wireless RF transceiver [1]. The large number of camera nodes forms a distributed system where the camera nodes are able to process image data locally (in-node processing) and to extract relevant information, to collaborate with other-cameras—even autonomously—on the application specific task, and to provide the system user with information-rich descriptions of the captured scene.

In the last years, several research projects produced prototypes of embedded vision platforms which may be deployed to build a visual WSN. Among the first experiences, Panoptes project [8] aimed at developing a scalable architecture for video sensor networking applications. The key features of Panoptes sensor are a relatively low-power and high-quality video capturing device, a prioritizing buffer management algorithm to save power and a bit-mapping algorithm for the efficient querying and retrieval of video data. Nevertheless the size of the sensor, its power consumption, its relatively high computational power and storage capabilities makes Panoptes sensor more akin to smart high-level cameras than to untethered low-power low-fidelity sensors.

The Cyclops project [9] provided another representative smart camera for sensor networks. The camera nodes is equipped with a low-performance ATmega128 8-bit RISC microcontroller. From the storage memory point of view the system is very constrained, with 128 KB of FLASH program memory and only 4 KB of SRAM data memory. The CMOS sensor supports three image formats of 8-bit monochrome, 24-bit RGB color, and 16-bit YCbCr color at CIF resolution (352 × 288). In the Cyclops board, the camera module contains a complete image processing pipeline for performing demosaicing, image size scaling, color correction, tone correction and color space conversion.

In the MeshEye project [10], an energy-efficient smart camera mote architecture was designed, mainly with intelligent surveillance as target application. MeshEye mote has an interesting special vision system based on a stereo configuration of two low-resolution low-power cameras, coupled with a high resolution color camera. In particular, the stereo-vision system continuously determines position, range, and size of moving objects entering its fields of view. This information triggers the color camera to acquire the high-resolution image subwindow containing the object of interest, which can then be efficiently processed.

Another interesting example of low-cost embedded vision system is represented by the CMUcam3 [11], developed at the Carnegie Mellon University. More precisely, the CMUcam3 is the third generation of the CMUcam series, which has been specially-designed to provide an open-source, flexible and easy development platform with robotics and surveillance as target applications. The hardware platform is more powerful with respect to its predecessors and may be used to equip low-cost embedded system with simple vision capabilities, so as to obtain smart sensors. The hardware platform is constituted by a CMOS camera, an ARM7 processor and a slot for MMC cards. Standard