Multimodal Video Analysis on Self-Powered Resource-Limited Wireless Smart Camera

Michele Magno, *Member, IEEE*, Federico Tombari, *Member, IEEE*, Davide Brunelli, *Member, IEEE*, Luigi Di Stefano, *Member, IEEE*, and Luca Benini, *Fellow, IEEE*

***Abstract*—Surveillance is one of the most promising applications for wireless sensor networks, stimulated by a confluence of simultaneous advances in key disciplines: computer vision, image sensors, embedded computing, energy harvesting, and sensor networks. However, computer vision typically requires notable amounts of computing performance, a considerable memory footprint and high power consumption. Thus, wireless smart cameras pose a challenge to current hardware capabilities in terms of low-power consumption and high imaging performance. For this reason, wireless surveillance systems still require considerable amount of research in different areas such as mote architectures, video processing algorithms, power management, energy harvesting and distributed engine. In this paper, we introduce a multimodal wireless smart camera equipped with a pyroelectric infrared sensor and solar energy harvester. The aim of this work is to achieve the following goals: 1) combining local processing, low power hardware design, power management and energy harvesting to develop a low-power, low-cost, power-aware, and self-sustainable wireless video sensor node for video processing on board; 2) develop an energy efficient smart camera with high accuracy abandoned/removed object detection capability. The efficiency of our approach is demonstrated by experimental results in terms of power consumption and video processing accuracy as well as in terms of self-sustainability. Finally, simulation results show how perpetual work can be achieved in an outdoor scenario within a typical video surveillance application dealing with abandoned/removed object detection.**

***Index Terms*—Embedded smart camera, energy efficient, infrared sensor, multimodal video surveillance system, wireless sensor network.**

I. Introduction

RECENT advances in micro-electromechanical systems, embedded computing, and low-power radio communication technology have sparked the advent of massively distributed wireless sensor networks (WSNs). The WSNs consist of large number of low-cost, low-power sensor nodes, which collect and disseminate environmental data. These sensor nodes are aimed at working within several scenarios, including surveillance, target acquisition, situation awareness, and chemical, biological, radiological, and nuclear early warning. To support these capabilities, it is now necessary to develop new architectures and design concepts that offer multimodal sensing without sacrificing the attractive low size, weight, and power capability offered by the conventional motes. The key advantage of WSNs is the ability to bridge the gap between physical and logical world by collecting and sending useful information to devices that have the computational resources to process it. WSNs, appropriately applied to dangerous tasks, can greatly decrease their risk, or even avoid the need of manpower for safety control.

Within this context, applications that exploit low-power video wireless networks (LP-VWN) consisting of networks of low-cost video sensors connected by low-rate wireless channels and constrained by low-power budget, have gained increasing attention [1]. In fact, a huge number of applications in surveillance, health care, environmental monitoring, and entertainment find interest in LP-VWN. Typical applications are in the domain of object detection, recognition, and tracking and a very challenging task is designing distributed video systems within the tight power budget typical of mobile devices and wireless sensor networks. These tasks could be performed after the acquisition of a continuous video stream on a power unconstrained base station. However, this approach would be extremely energy and bandwidth inefficient, difficult to implement on stand-alone mobile embedded systems and ultimately not scalable in a network. Clearly, nothing should be done from the point of view of data transmission if the target object/event is not detected. Even in presence of the target object/event, only some very limited amount of information may be transmitted, such as the number of interest objects, their size, position, trajectory, etc. In terms of computing power, using smart cameras reduces the processing load of the central processing units by means of the execution of low-level image processing tasks within the camera platform and before data transmission to the host system. This way, the amount of data needing to be transmitted is radically reduced since, instead of sending the whole image contents to the host system, only some specific, postprocessed information is sent. Furthermore, transmitted data is more pertinent than the raw pixel flow, meaning that received data can be promptly used by the central processing units, without the need for running time-consuming tasks.

Energy is one of the scarcest resources in wireless sensor networks. This key issue is more critical for power hungry applications such as video processing. To enhance vision

Manuscript received January 31, 2013; revised March 22, 2013; accepted March 26, 2013. Date of publication April 23, 2013; date of current version June 07, 2013. This work was supported in part by the ENIAC Joint Undertaking under Grant 120214 (END project), in part by the projects GENESI (Grant 257916) and 3ENCULT (Grant 260262) funded by the EU 7th Framework Program, and in part by the Autonomous Province of Trento with "EnerViS–Energy Autonomous Low Power Vision System" project. This paper was recommended by Guest Editor A. Prati.

M. Magno and L. Benini are with the Department of Electrical, Electronic and Information Engineering, University of Bologna, 40136 Bologna, Italy (e-mail: michele.magno@unibo.it; luca.benini@unibo.it).

F. Tombari and L. Di Stefano are with the Department of Computer Science and Engineering, University of Bologna, 40124 Bologna, Italy (e-mail: federico.tombari@unibo.it; luigi.distefano@unibo.it).

D. Brunelli is with DII, University of Trento, 38100 Povo, Trento, Italy (e-mail: davide.brunelli@unitn.it).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JETCAS.2013.2256833

sensor networks, two successful strategies can be adopted: 1) exploiting alternative power sources, which increase the autonomy of the nodes; 2) exploring multi-modal sensor integration, which can save on-board power consumption. Recently, several researchers have proposed alternative power sources and energy harvesting techniques to replenish energy buffers like batteries or super capacitors by extracting and converting power from the surrounding environments [18]–[20]. Energy harvesting technologies are used to collect energy from ambient sources. An energy harvesting device converts this energy into electric energy which is stored in the energy storage device of the sensor node. In particular, photovoltaic (PV) harvesters appear as the most promising for enabling perpetual operation of WSNs [2], [3].

On the application side, the main focus of this work is on video surveillance, processing images directly on board, only when required. So this approach is completely orthogonal to the CCTV approach which sends out or record a huge amount of video data that is then processed offline. Specifically, our goal is to develop a video surveillance system that, through the analysis of image data acquired from the video sensor, performs automatic and reliable detection of abandoned and removed objects in the monitored scene. Due to the on-board video processing is possible automatically detect if some object is abandoned of removed from the monitored field of view and give alarms via wireless radio. Typical scenarios for the proposed system are, e.g., detection of unattended packages in a railway station or in an airport, or real-time traffic monitoring and surveillance [4], [6] as well as detection of stolen objects in a museum [5]. In particular, automatic abandoned object detection in small public indoor environments, such as toilets and lavatories, is particularly suited to the characteristics of the proposed device (low-cost, low-power, limited spatial sensor scope). Several approaches have recently addressed this specific task [4]–[15], [16], typically relying on off-the-shelf computing platforms such as personal computers or wired/wireless commercial cameras. However, these systems are typically power hungry, thus not suitable to work within wireless sensor network applications, where power consumption is a critical issue. In fact, the design of such video analysis algorithms on a smart camera characterized by low power consumption, low processing capability, and small size is challenging due to the limited amount of available resources. It is worth pointing out that none of the previously mentioned approaches is based on a similar embedded architecture. Moreover, we believe that combining local processing, low power consumption, and power management, energy harvesting, self suitability, and distributed intelligence is a key challenge to make video surveillance based on wireless sensor networks a reality. Thanks to these features the LP-VWN can compete or improve PC-based and commercial USB/LAN cameras video systems and applications.

In this paper, we present a multi-modal video surveillance system based on a complementary metal–oxide–semiconductor (CMOS) video sensor and a pyroelectric ("passive") infrared (PIR) sensor characterized by low power consumption and low cost to be used as a node in a WSN. Moreover, the node is equipped with a solar panel that we later show to enable perpetual operation. We propose a solution integrated into a stand-alone camera with embedded video processing capabilities and wireless communication. The proposed application relies on an advanced video analysis framework that, based on the same low-cost and low-power architecture, is able to detect events such as abandoned or removed objects. The PIR sensor is integrated with the video processing module, since it appropriately triggers the video analysis module based on the absence/presence of people in the scene. This provides two main benefits. The first one concerns the robustness of the video analysis algorithm, since, as it will be shown more in details in the following, it helps reducing false positives due to occlusions or moving objects. The second one concerns power consumption: by limiting the activity of the video analysis module when this is not needed, there is a notable reduction of the overall power consumption of the system, as shown in [17]. In fact, in the aforementioned scenario of detection of abandoned/removed objects with a camera sensor network, most of the time the surveyed area is empty and the network should no longer monitor continuously the scene because there is nothing to detect. When an event is detected by the PIR, the network can be switched on and begin video processing once again. Although the camera is low power, it can last only few hours in continuous mode, however thanks to the energy management policy that uses a sleep and wake-up strategy for energy conservation together with a PV harvester, the video node can work perpetually in the proposed scenario (with 50 events per hour).

The remainder of the paper is organized as follows. Related work is discussed in Section II, while in Section III we present the system architecture focusing on the constraints of energy budget, memory and computational capability offered by an ARM-based solution. Section IV describes the image processing algorithm used for the detection of abandoned/removed objects, discussing the constraints met and the implementation details developed to deal with this limited platform. Experimental measurements and achieved performance are the focus of Section V. Finally, Section VI draws conclusions.

II. Related Work

A video sensor node in a wide area sensor network is a node capable of performing on-board video processing, and communicating the information over a self-organizing and fault-tolerant wireless network. Commercial CCTV and PC-based surveillance systems are not suitable for wireless sensor network surveillance application, and are orthogonal to our approach. In fact, the CCTV are not processing the data on line and are wired for power consumption, while PC-based and commercial cameras rely on high power consumption, which is not suitable for the WSN. For this reason, this section does not review these kinds of solutions.

Recently, several platforms and commercial cameras with similar goals have been developed within a sensor [21]–[29].

We can classify these approaches in three categories:

- low-cost nodes with wired interface (e.g., commercial USB/LAN camera, or the node designed by Corely *et al.* at CMU [22]);
- wireless nodes with significant power consumption (e.g., commercial wireless LAN or the Panoptes nodes designed by Feng *et al.* [25]);

- application-specific single ultra-low power single chip solution (e.g., the chip designed by Zhang *et al.* [24]).

Being wired, the nodes in the first category obviously do not satisfy the basic requirement of being wireless, thus usually these cameras send the data to a remote host to be processed, although novel commercial cameras have the possibility to do it on-board they still send the results by means of a wire. Instead, nodes in the second category consume roughly $10\times$ more power than a typical first generation wireless sensor node and for this reason they are not suitable for wireless sensor network which should work for several months. Finally, single-chip solutions have extremely low-power consumption, but they are neither programmable nor configurable in field. One important common point among current video wireless nodes belonging to the first and second category is that the digital signal processing subsystem is the main power bottleneck. This is due to the fact that the high data rate of CMOS image sensor imposes the selection of fast processors and memories with high power consumption. Hence, the main open challenge in this area is to synergistically develop algorithms and architectures for energy-efficient image processing without giving up the flexibility of in-field configuration.

The academic literature provides more similar approaches in terms of low-power cameras and wireless sensor application. In [30] sensor nodes equipped with PIR, acoustic and magnetic sensors have been deployed in order to achieve adjustable sensitivity, stealthiness and effectiveness in a distributed military surveillance applications. To balance privacy and security in surveillance applications, networks of infrared sensors (IR) and cameras are employed also in [31]. Cameras are used in public areas while networks of IR detectors are deployed in private areas. The system processes the data from IR sensors in order to detect an event of interest in the private area to identify the author by correlation with the images grabbed in the public area. Energy conservation through limiting the sensing to a small part of the network was also considered in [32]–[35]. In [32], the activation pattern is swept across the network. Both schemes assume a simple topology and do not handle sensing holes due to the sleeping time of the sensors. In fact in this previous work the sleeping time is fixed and the sensor could sleep also when it is need have data acquisition. In [33], the activation pattern follows a user-defined path through the sensor network as a sentry. In [36], a sleep/wakeup strategy in solar-powered wireless sensor is presented. Our work incorporates many similar ideas to the ones mentioned above however in contrast to the other works we present a combination of video sensor with other low-cost and low-level sensors, which are used mainly for triggering the camera at the right time and not to promote a reduction of the system energy requirements. In fact the PIR sensor can be used as an ultra low power wake-up trigger to reduce the power consumption when the video processing is not needed. Moreover, data from PIR sensors can be used from the application to understand when the best moment for starting an image acquisition occurs. This will bring both a reduction of false positives and a reduction of the number of frame being processed, thus also a reduction of power consumption. Finally, we equipped the node with a solar panel to recharge the batteries and achieve perpetual work.

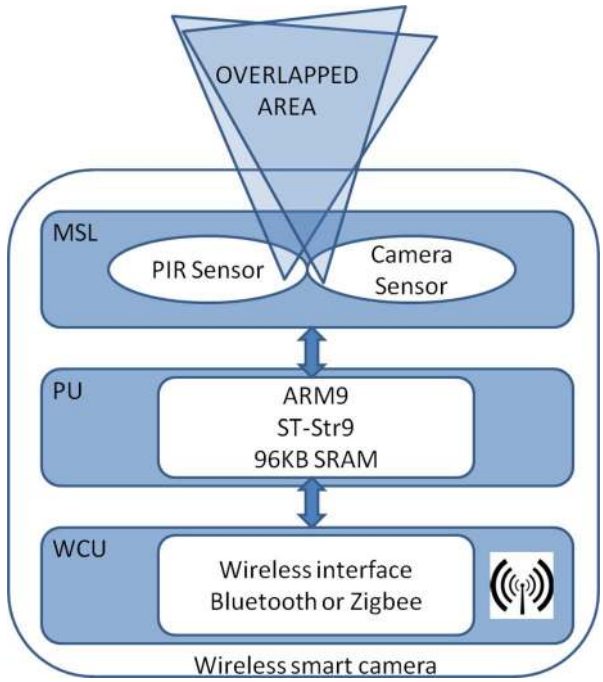

Fig. 1. Multisensor layer (MSL) includes both camera sensor and the PIR sensor; processing unit (PU) is the core of the node and include an ARM9 microcontroller with SRAM. Wireless communication unit (WCU) provides the interface with a Bluetooth or Zigbee wireless board.

III. System Architecture

Fig. 1 shows the smart camera structure. A smart camera is a camera that can do on-board processing instead of transmitting all video data to a central controller. The choice of a smart camera is motivated by the notably reduced power consumption required by processing on-board compared to a solution relying on transmitting raw data through a wireless interface. The wireless smart camera consists of a camera sensor, an embedded processor (ARM9 STR912 from STM), a CMOS video sensor (VS6624 from STM), a pyroelectric infrared sensor used as trigger and a wireless communication module, being supported through a suitable interface which can host either a ZigBee or Bluetooth compliant transceiver. The whole system is designed to achieve low power consumption. Each device provides a power saving mode to reduce consumption when not in use. The smart camera is connected to a host (gateway or central PC) to send alarms or relevant image content; moreover it can receive wireless messages from the host so as to modify its settings.

A. Processing Core

Choosing the suitable target hardware for a smart camera processor is an important issue. Due to the very large amount of data involved in image processing and computer vision tasks and the aforementioned constraints on power consumption, we have chosen an STR912F microprocessor from STMicroelectronics. It is based on an ARM966E 16/32-bit RISC architecture working up to 96 MHz operating frequency. On the chip there are 96 KB SRAM and several interfaces including Ethernet, USB, I2C, and UART. Since one main goal of the system is low power operation, most of the peripheral interfaces have been

disabled in our design. One of the major features of this MPSoC is the configurable and flexible power management control that allows the user choosing the best power option that fits the current application. Thanks to this feature, power consumption can be dynamically managed by firmware and hardware to match the system requirements adjusting the operating frequency or changing the processor state. STR912F supports three global power control modes: run, idle, and sleep. The typical current consumption for this microcontroller is about 1.7 mA/MHz in Run mode and only a few mA in sleep mode, which is an attracting feature for wireless sensor network design where the power consumption is a main constraint. To capture images from the camera and processing data for image classification we employ a high-speed logic interface provided from STR912F.

B. CMOS Image Sensor

To achieve low power consumption we chose the VS6624 CMOS imager from STMicroelectronics as the video device. The image sensor provides full SXGA (1280 × 1024) resolution at 15 frames per second, or VGA (640 × 480) resolution at 30 frames per second. A very important feature is the power consumption, which is just 120 mW when active (2.8 V and 12 Mhz frequency), while it goes down to 23 mW when it switches to standby. The CMOS camera can be programmed and controlled via internal registers using I^2C serial interface. It supports several output formats, however most video processing algorithms use grayscale image, thus we adopt 8-bit grayscale images with YCbCr 4:0:0 format. Although it supports SXGA resolution, due to limited size of the internal SRAM only little parts of the images can be processed, typically 160 × 120 pixels, but the node can use external RAM extending the memory capability. This resolution is enough to perform our image processing algorithm, and to save time and energy for storing and processing data.

C. Pyroelectric Infrared Sensors

Pyroelectricity is the electrical response of a polar, dielectric material (crystal, ceramic, or polymer) to a change in its temperature. The basic model of a pyroelectric element is a planar capacitor whose charge Q changes according to $\Delta Q = Ap\Delta T$, where A is the area of the element, p is the pyroelectric coefficient of the material, and T is the temperature [37]. Using electrodes, this charge can be detected as a current flowing through an external circuit such that $I = Ap\,dT/dt$ [38]. The incident radiation causes the change in temperature of an absorbing structure that is designed to maximize the ΔT at the required wavelength.

Commercial PIR detectors typically include two sensitive elements placed in series with opposite polarization (Fig. 2). Such a configuration makes the sensor immune to slow changes in background temperature. PIR sensors are used in conjunction with Fresnel lenses. The aim of the lenses is both to shape the field of view of the detector and modulate incident radiation by optically dividing the area to be protected into a number of separate cones. PIR sensors are largely used in modern alarm systems to detect presence of people and provide a simple, but reliable, digital presence/absence signal, being reliable and having low prices and low power consumption.

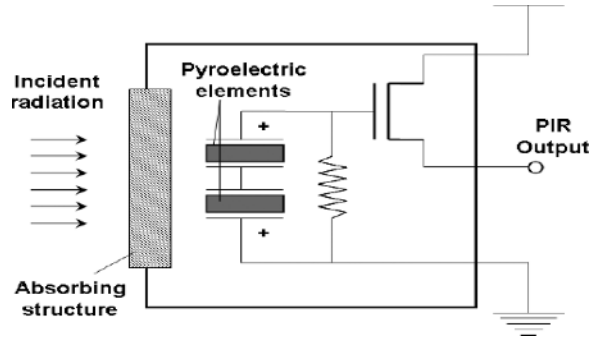

Fig. 2. Typical commercial PIR sensor. Two sensitive elements are placed in series with opposite polarization in order to gain background temperature immunity.

TABLE I
CHARACTERISTIC OF MURATA IRA E710 PIR DETECTOR

Feature	Unit Symbol	Value
Responsivity (500K, 1Hz, 1Hz)	**MVPP**	**4.3**
Field of View	°	90
Spectral Response	m	5-14
Elements	mm	2 x (2 x 1)
Supply Voltage	V	2 to 15
Operating temperature	°C	-40 to 70

PIR sensor conditioning circuits: Although the PIR analog output is theoretically more informative for the following video analysis stage, in this work we built the circuit to have a digital output, given its increased robustness. The detector used for our smart camera is Murata IRA E710, which presents the characteristic shown in Table I and the adopted Fresnel lenses are IML-0635 from Murata. The bias output voltage is $V_{dd}/2$ and the output signal of the sensor needs to be amplified several hundreds of times and filtered in order to be processed by a digital system. Thus we built a two-stage amplifier which achieves a total amplification of about 1500 times and filters the signal between 0, 5, and 12 Hz.

In addition to the amplifier, we designed a trigger with adjustable threshold. The schematic of the circuit is presented in Fig. 3. As long as the PIR output does not exceed lower and upper thresholds set by the two resistors R1 and R2, and the value of the digital potentiometer, the trigger signal is kept at V_{dd} by the pull-up resistor. Passing one of the two thresholds, an OPAMP pulls the trigger signal to GND; generating an interrupt for the microcontroller of the smart camera. Using an identical resistance for R1 and R2, the two thresholds are symmetrical to $V_{dd}/2$, and their reciprocal distance increases with the resistance of the digital potentiometer. In this manner we can program dynamically and run time the sensitivity of the wake-up signal for the camera according to the constraints placed by the status of the system. The I^2C interface of the STR912F is connected to the digital potentiometer to dynamically change the threshold of the trigger. It is thus possible to increase or decrease the characteristics of the field of view of the PIR sensor.

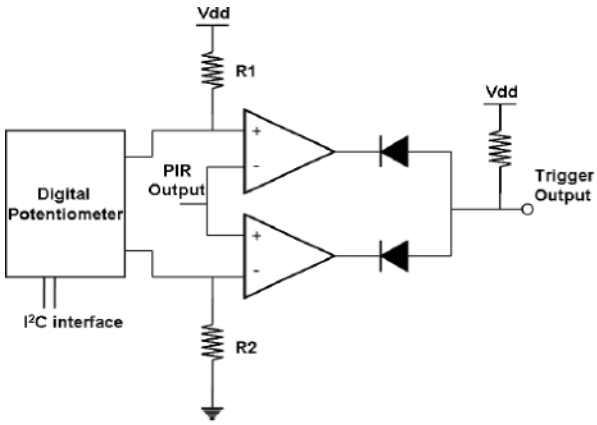

Fig. 3. Schematics for trigger generation using PIR output signal. I²C interface is used to adapt the sensitivity of digital output (trigger output) of the PIR sensor. The PIR output is analog signal of the PIR. The comparators and the diodes generate the digital output.

D. Wireless Transceiver

The smart camera hardware and software interface is built to host different wireless standards used in the wireless sensor network community, such as ZigBee and Bluetooth or proprietary protocols. However, in this work we use WT12 from Bluegiga, a highly integrated Bluetooth module, containing all the necessary elements from Bluetooth radio to the antenna. Performance and measurements discussed in this paper henceforth refer to the version with Bluetooth capability.

E. The Energy Harvesting Unit

Energy harvesting is a low cost and effective operation, in term of energy harnessed, device size, and efficiency. One of the primary issues that need to be tackled is minimizing the power consumed by the harvester itself. The less the power required by the circuit, the faster the growth of the harvested energy in the accumulator. Solar harvesting is mainly used for sensor nodes deployed outdoor and several circuits have been proposed to increase the autonomy of embedded systems. The *I–V* characteristic of a *PV* module is given by the following equation:

$$I_o = I_g - I_{\text{sat}} \left\{ \exp\left(\frac{q}{AKT} * (V_o + I_o R_s) \right) - 1 \right\} \qquad (1)$$

where Ig is the generated current, I_{sat} is the reverse saturation current, q is the electronic charge, A is a dimensional factor, K is the Boltzmann constant, T the temperature in degree Kelvin, and R_s the series resistance of the cell. The internal shunt resistance is neglected in this model. The plot of the PV module adopted in our solar harvester is shown in Fig. 4.

For our smart camera we used the SOLAREX MSX-005F solar cell, which presents the characteristics shown in Table II. Furthermore we used a COTS boost converter and the Maxim ic MAX1551 to recharge a 2000 mA/h Li-Ion battery. To increase the energy efficiency of solar harvester the smart camera uses the device developed and modeled in [40]–[42] which includes a maximum power point tracking (MPPT) algorithm and can perform high efficiency around 90%. Furthermore, we used a COTS boost converter and the Maxim IC max1551 to recharge a 2000 mA/h Li-Ion battery.

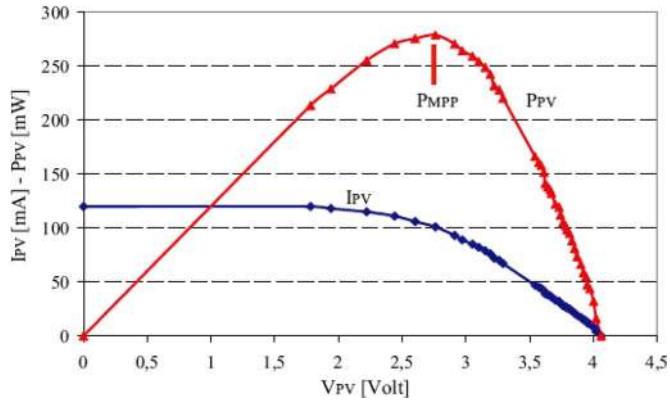

Fig. 4. Characteristic of the photovoltaic module.

TABLE II
CHARACTERISTIC OF SOLAREX MSX-005F SOLAR CELL

Feature	Parameter	*Value*
Nominal open circuit voltage VOC	V	4.6
Nominal voltage @ maximum power point VMPP	V	3.3
Nominal short circuit current ICC	mA	160
Nominal current @ maximum power point IMPP	mA	150
Lead Length	mm	750
Power Rating	mW	500
Size-Length	mm	114
Size-Width	mm	66.8

IV. VIDEO ANALYSIS ALGORITHM

As described, the primary goal of the proposed wireless sensor node is to check for "abandoned/removed object" events in the monitored scene as soon as the status of the system goes from "presence of motion" to "absence of motion." We have already explained how the use of the PIR sensor allows for sensing the presence of movements in the monitored scene so as to detect the occurrence of this specific event. Hence, we propose here a video analysis algorithm aimed at performing the abandoned/removed object detection suited to the traits of our system. Overall, this algorithm first detects stationary changes in each frame acquired by the video sensor, then classifies the detected changes between either "removed" or "abandoned."

Due to the constraints linked to the low-power and embedded traits of our platform, the camera does not acquire a video sequence, and instead it only captures a snapshot of the current scene, which is then compared with the previous frame (i.e., the one that was acquired before the last motion took place). This requires a small memory footprint, compatible with the RAM memory available aboard the platform, and allows for reducing the transfer time between the memory and the processor. In addition, it is worth pointing out that, due to the aforementioned reasons, all stages of the proposed video analysis algorithm have to be particularly efficient to deal with the adopted low-frequency processor, and need to avoid the use of floating point instructions given the absence of a floating point unit (FPU) within the architecture.

Several recent approaches are present in literature for the task of motion detection in video sequences [43]–[45]. Differently

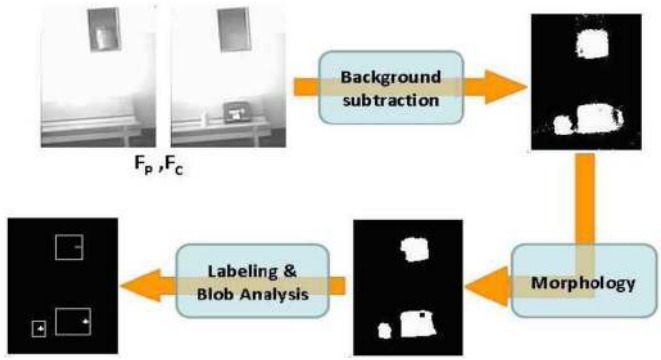

Fig. 5. Flow diagram of the proposed video analysis algorithm.

from their case though, here the video analysis is more challenging due to the absence of a full set of frames to be processed. In particular, it is hard to determine the detected changes only based on the single acquired snapshot. Nevertheless, a working assumption brought in by the joint use of the PIR sensor and the camera that simplifies the video analysis task is represented by the fact that all visible changes appearing in the scene in absence of movements can be considered stationary changes, thus possible instances of removed or abandoned objects.

The proposed video analysis algorithm is based on three stages. During the first stage, the current image is compared to the stored one in a *background subtraction* manner yielding a change mask aimed at highlighting visible changes in the scene appearance. Following this step, an image processing stage analyzes the change mask to filter changes due to the presence of nuisances in the data. In particular, this step relies on labeling the connected components of the image [or region-of-interest (ROI)] and eliminating those whose area is below a given threshold. Finally, a higher-level vision algorithm performs a blob analysis aimed at classifying each ROI between abandoned and removed object. Fig. 5 shows a flow diagram of the algorithm.

A. Background Subtraction

The first stage of the video analysis algorithm aims at obtaining a change mask highlighting the visible changes (i.e., changes in pixel intensity) between the current frame F_C and the previously acquired one F_P In our approach we adopt a typical *background subtraction* algorithm, by interpreting the frame F_P as the current scene background, since referred to the appearance of the scene previously to the last occurred motion. Hence, we compare each pixel at coordinates (x, y) in F_C with its homologous on F_P by means of a function aimed at measuring the similarity between the two image points. In our case, since a static camera is adopted, homologous pixels have identical image coordinates

$$S(x, y) = f(F_c(x, y), F_p(x, y)). \tag{2}$$

The standard and simplest approach would simply compute this similarity as a function of the absolute difference of the two pixel intensities. Although simple, this approach is particularly sensitive to the nuisances typically present in image data under real working conditions such as e.g., camera noise and light changes. This would easily lead to misinterpreting nuisances as structural changes, yielding an abundance of false positives in the change mask. Therefore, we adopt a more robust approach that relies on taking into account all pixel intensities belonging to a given local support of the point at coordinates (x, y) on F_c and F_p. In particular, we compare the two squared pixel windows of size r and centered at coordinates (x, y) on the two frames by means of the normalized cross-correlation (NCC) [46]

$$\mathrm{NCC}(x, y) = \frac{F_C(x, y) \circ F_p(x, y)}{\|F_c(x, y)\|_2 \cdot \|F_p(x, y)\|_2} \tag{3}$$

where the numerator is the dot product between F_c and F_p

$$F_c(x, y) \circ F_p(x, y) = \sum_{i=-r}^{r} \sum_{j=-r}^{r} F_c(x + i, y + j) \cdot F_p(x + i, y + j) \tag{4}$$

and the two terms at the denominator represent the L_2 norms of $F_c(x, y)$ and $F_p(x, y)$, respectively

$$\|F_c(x, y)\|_2 = \sqrt{\sum_{i=-r}^{r} \sum_{j=-r}^{r} F_c^2(x + i, y + j)} \tag{5}$$

$$\|F_p(x, y)\|_2 = \sqrt{\sum_{i=-r}^{r} \sum_{j=-r}^{r} F_p^2(x + i, y + j)}. \tag{6}$$

The use of the NCC is advantageous on one side since it is invariant to linear photometric transformations between the two corresponding windows on F_c and F_p, on the other side because its computation is still particularly simple and efficient. Since, as already mentioned, there is no FPU in our target architecture, to perform the square root and division operations in (3) a fixed-point square root function for ARM and an integer division have been implemented. Successively, a threshold is applied on the NCC score at each pixel location, this yielding the binary change mask, C, which highlights those parts of F_C which have been subject to a change with respect to F_P

$$C(x, y) = \begin{cases} \text{changed}, & \mathrm{NCC}(x, y) < \tau_N \\ \text{unchanged}, & \text{otherwise} \end{cases}. \tag{7}$$

Other measures do exist that tolerate more general classes of transformations than the linear, e.g., *order preserving* transformations that assume that nuisances cannot change the order between neighboring intensities. Examples of such approaches include the rank and census transforms [47], and more recently [46], [48], [49]. Moreover, very effective methods based on the order preserving assumption and a probabilistic formulation of the binary classification problem given by (7) have been proposed [50]–[52]. Although allowing for improved robustness to typical nuisance factors, these methods are significantly more expensive under both the computational and memory requirement point of view so that they definitely do not match the limited computational resources available in our architecture.

A typical drawback of background subtraction approaches relying on a spatial support, such as our NCC-based method, is that the segmentation of the foreground in the change mask

becomes less accurate along the borders of the objects. In particular, there's a typical *fattening* effect, i.e., the object appears bigger since its borders are increased by a number of pixels proportional to r. To deal with this effect, a simple binary morphology operator of erosion is applied to the change mask as many times as the chosen value of r. The application of this operator has also the beneficial effect of reducing the presence of noise in the change mask.

In order to have a refreshed background and to avoid background subtraction mistakes, the algorithm updates the background under the following condition: 1) at the end of the video processing algorithm the current frame becomes the new background; 2) after a fixed or adaptive (according with the firmware) time interval of inactivity, the camera wakes up and updates the background.

B. Labeling

Given the output of the background subtraction stage, a labeling algorithm is applied to identify the connected components present in the binary change mask. This operation allows for determining the number of separated objects that have been removed/abandoned in the current frame. More specifically, we adopt the approach proposed in [53], an efficient labeling algorithm for binary images which also has low memory requirements. In particular, the algorithm only requires two image scans and has a memory complexity of O(1).

Then, a bounding box is computed for each connected component, representing the squared ROI of that specific component. The ROI computation only needs one additional image scan and it has two main advantages. First, it allows compressing the information concerning the area of the image that was subject to a change: this is useful since the sensor only needs to send to the base station the two coordinate pairs defining the ROI, this allowing less bandwidth usage and higher transmission speed. Secondly, it allows the application of a postprocessing step based on the ROI shape aimed at eliminating spurious connected component originated by noise. In particular, all ROIs having one side smaller than a predefined threshold are eliminated from the further processing stage.

C. Blob Analysis

This last stage of the video analysis algorithm aims at classifying each ROI either as an object abandoned or removed. The main assumption on which the adopted classification algorithm relies on is that the presence of a foreground object tends to increase the number of edges in the image around the borders of the object compared to the background. Hence, if an object is abandoned on the background, the number of edges along the borders of the corresponding connected component on F_C should increase compared to F_P. Conversely, if an object is removed, then F_C should display a decreased number of edges along the borders of the connected component related to the original position of the object on F_P.

Hence, the approach relies on the estimation of the number of edges that appear on F_C along the borders of the connected component we wish to classify (Fig. 6). First of all, we detect all *contour* points within the ROI as those points that belong to the foreground and have at least one of their 8-connected neighbors

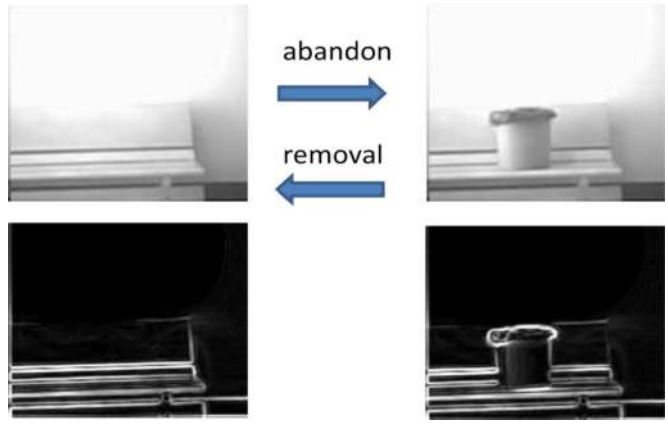

Fig. 6. Edge-wise, the presence of foreground objects in the scene which were not present in the original background model tends to create additional edges along the border of the object, vice versa, when an object is removed from the background, border edges will mostly disappear. This is clearly visible by comparing the two edge maps shown at the bottom, associated to their respective frame on top. Upon accurate determination of the object borders (performed in the previous background subtraction and morphology stages), the blob detection step is thus able to robustly classify the object as either abandoned or removed.

set as background. On each contour point at coordinates (x, y), we compute the horizontal and vertical derivatives Dx, Dy of point $F_C\ (x, y)$. Specifically, the derivatives are computed by means of the Sobel operator, since it provides for a robust and fast derivative estimation that requires integer operations only. Then, we estimate the magnitude of the gradient at (x, y) as

$$|G(x,y)| = \max(|D_x(x,y)|, |D_y(x,y)|). \tag{8}$$

A threshold, i.e., τ_E, is used to classify the contour point as being or not an edge in F_C. Then, the percentage of contour points associated with edges, N_{CE}, is computed as the ratio between the total number of contour points associated with an edge and the total number of contour points. This percentage value is then thresholded to yield the final classification of the ROI

$$\text{Class}(x,y) = \begin{cases} \text{removed}, & N_{CE}(x,y) < \tau_C \\ \text{abandoned}, & \text{otherwise} \end{cases}. \tag{9}$$

Ideally, to yield a symmetric behavior of the classification algorithm towards the two classes (i.e., no bias needs to be introduced towards one specific class) the value of τ_C should be set to 0.5. This approach is iterated on all ROIs found on F_C by the previous stage of the algorithm so as to yield a classification of all relevant changes over the current frame of the scene.

V. Experimental Results

A. Video Analysis Evaluation

We present here two experiments, referred to as Experiment 1 and 2, aimed at assessing the capability of the proposed algorithm to correctly detect static changes in the monitored scene, and classify them as either abandoned or removed objects. For both experiments, the same set of parameter values is used, which is reported in Table III. As for τ_N, the value used is

TABLE III
PARAMETER USED THROUGHOUT THE PROPOSED EXPERIMENTAL EVALUATION

Stage	Parameter	*Value*
Background Subtraction	τ_N	$1023*2^{-10}$
Blob Analysis	τ_E	*160*
Blob Analysis	τ_C	*-0.5*

to allow performing the comparison in (7) using integer operations: in fact, in our implementation the NCC value at each pixel is multiplied by 2^{10} (i.e., left-shifted of 10 positions) and compared with an integer threshold (equal to 1023 in our experiments). Also, all the results concern data that has been wholly acquired and processed online and onboard the architecture.

Experiment 1 concerns an indoor setup aimed at monitoring a sector of a hallway in a building. The camera was placed a few meters away from the wall and switched on. In order to achieve the best performance, the camera has to be firmly attached to the wall or placed over a vibration-free surface. In fact, although the camera is mounted on a mobile device, the image processing algorithm assumes the presence of a static camera. Meanwhile several real abandoned/removed object instances were generated. The processed images were sent directly to a host PC via Bluetooth. We have analyzed three continuous footages where, each time the camera was switched on by the PIR sensor, the current frame was processed by the proposed algorithm. We have collected a total of 64 frames, with each frame containing at times even multiple abandoned/removed objects, for a total of 78 abandoned/removed objects. Also, a subset of frames (i.e., 11) does not contain any case of object removal/abandon. As for the background subtraction algorithm, out of 78 ROIs, 77 ROIs have been correctly detected by the algorithm, yielding only one false negative and four false positives, for a correct segmentation rate of 98.7%. These 77 ROIs, fed to the classification stage, were then correctly classified in 73 cases as abandoned/removed object, yielding to a correct classification rate of 94.8%.

Fig. 7 reports some qualitative examples of correct segmentation and classification yielded by the proposed algorithm on the dataset of Experiment 1. In the figure, seven examples are reported (indicated by letters a–f), where for each example the two frames F_P, F_C and the final output of the algorithm are shown. As for the output, each detected ROI is indicated by a white bounding box, while the classification result is encoded as a symbol shown inside each detected bounding box: the "+" symbol stands for abandoned object (as if it is added to the scene) while the "−" symbol stands for removed object (as if it is subtracted from the scene). As it can be seen from the figure, the proposed algorithm is able to achieve correct segmentation and classification also in presence of very small, partially transparent objects (cases e and f) and with multiple ROIs (cases a–f).

Fig. 8 reports some cases where our algorithm produced wrong results, by showing for each case F_P, F_C and the final output of the algorithm, as previously done in Fig. 7. In particular, it shows the only false negative (case c) and three out of the four reported false positives (case b, d, e) yielded by the background subtraction stage. In addition, it shows two

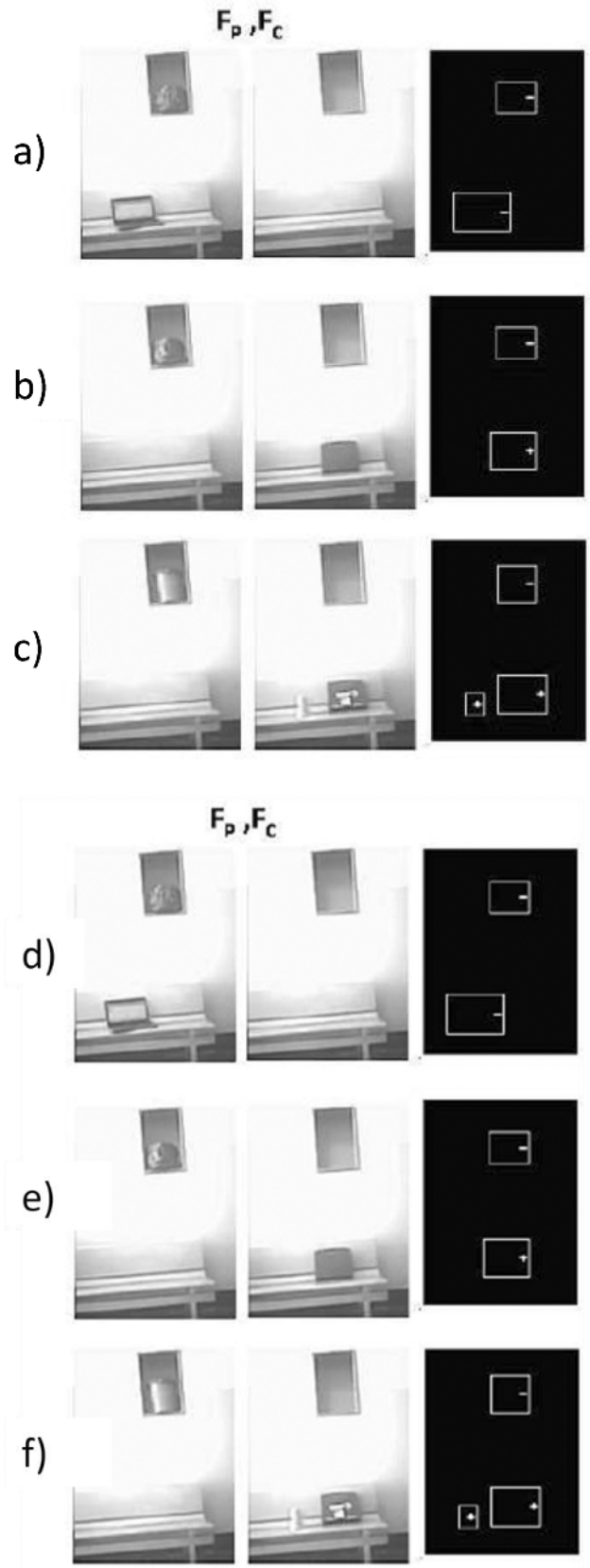

Fig. 7. Correct segmentation and classification examples yielded by the proposed algorithm on the dataset of Experiment 1.

of the four classification errors produced by the classification stage (case a–f). As it can be seen, the majority of wrong

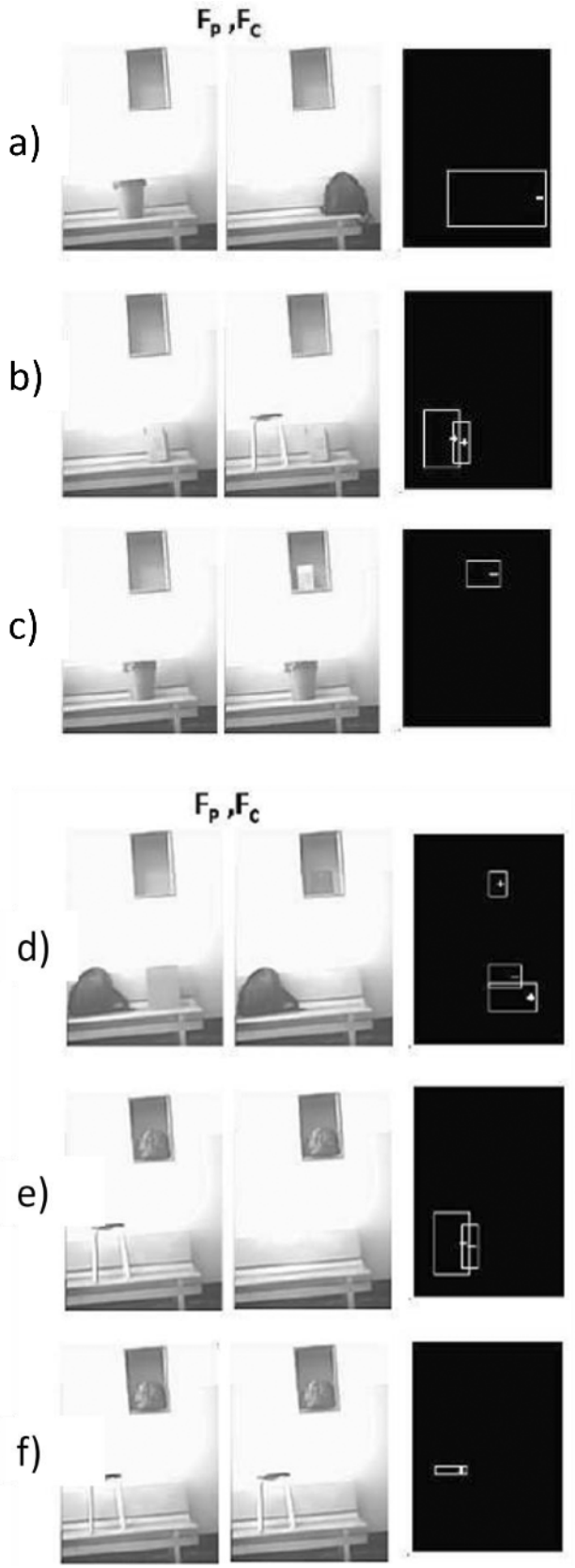

Fig. 8. Examples of errors yielded by the proposed algorithm on the dataset of Experiment 1. b–e: Errors due to background subtraction: false negative (c) and false positives (b, d, e). a, f: Classification errors of the Blob Analysis stage.

segmentations are due to splitting of the ROI into two sub-ROIs (case b–d). The other wrong segmentation (case a) is due to the too high proximity of the two objects, which are thus merged into the same ROI. As for classification, apart from the errors arising due to a wrong segmentation, we can note the error due to the very small removed object in case f.

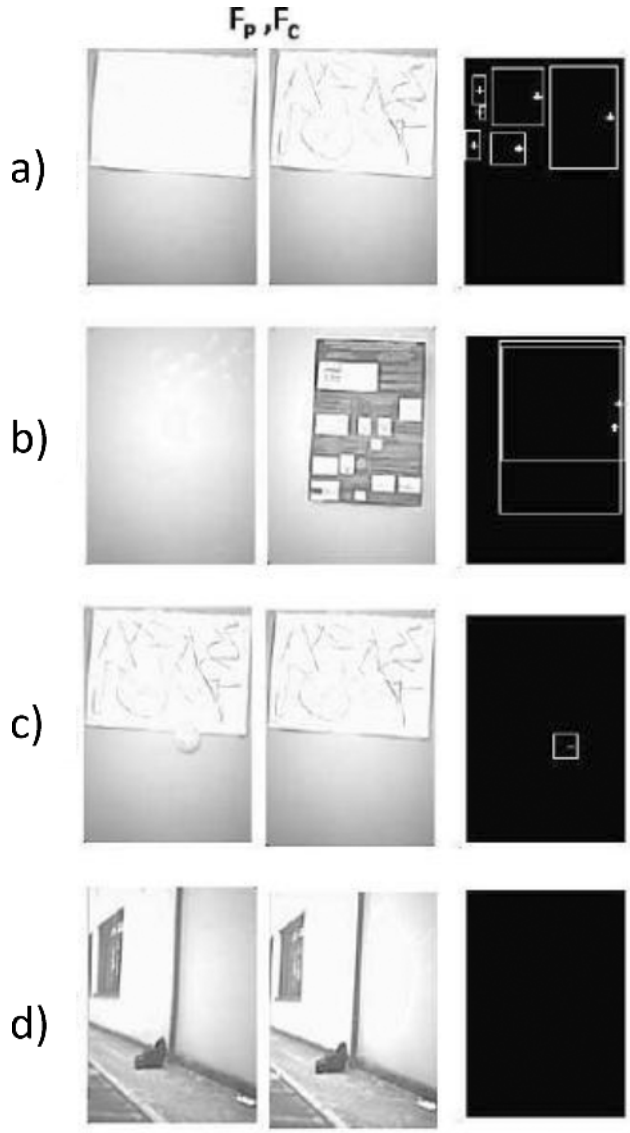

Fig. 9. Performance of the proposed video analysis algorithm in different outdoor circumstances: graffiti (a), object abandon (b), object removal (c), illumination change (d).

In Experiment 2, we present a qualitative evaluation of the proposed algorithm in an outdoor environment under sunny conditions, which concerns specifically the case of monitoring a wall from undesired events such as graffiti, stealing or unauthorized poststicking. In this case the distance from the wall was around 5 m in the case (a–c) Figs. 9 and 10 and 7 m in (d) Figs. 9 and 10. More specifically, and as it is shown in Figs. 9 and 10, our system has been tested against events such as graffiti (case a), wall affixtures (case b), object removal [case c and Fig. 10(d)]. Finally a situation where only illumination changes occur in the scene is shown in Fig. 9(d). In this test the camera was left at 1000 in morning and the video processing were performed 2 h later with different light. As it can be seen, our algorithm is capable of robustly handling all the

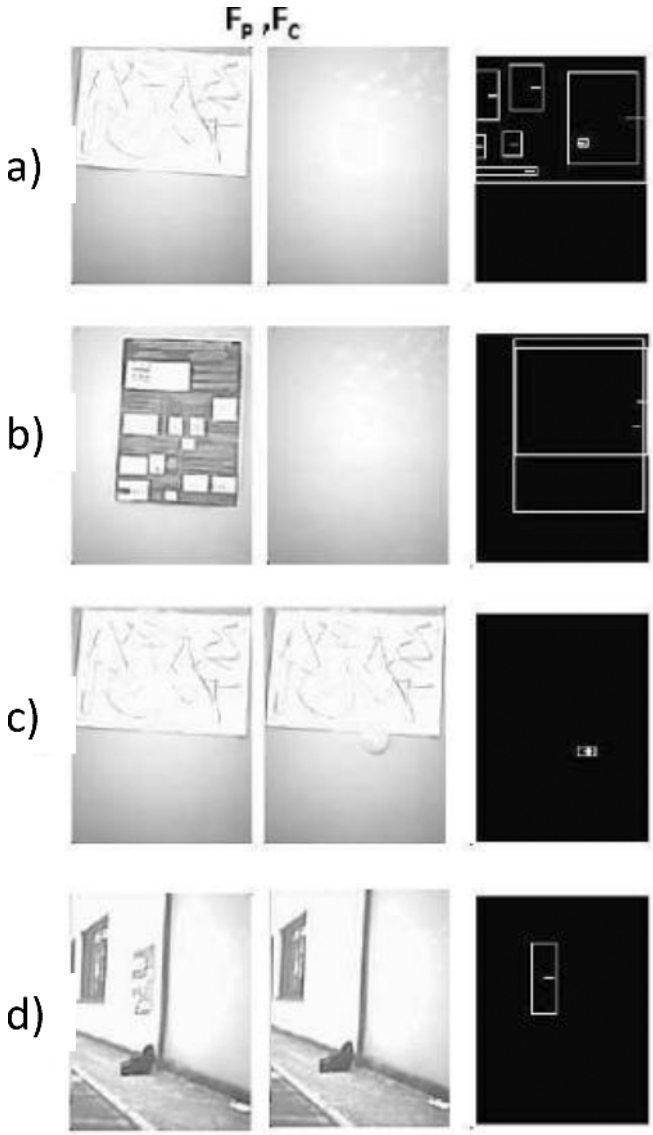

Fig. 10. Performance of the proposed video analysis algorithm in different outdoor circumstances: object removal (a, b, d), object abandon (c).

aforementioned cases, yielding correct classification also when the object ROI is particularly small (e.g., case c of both figures and d Fig. 10, and with each single writings appearing in the graffiti situation concerning case a) and despite the nonoptimal illumination conditions (e.g., the notable saturation present in the data due to the external bright sunlight).

B. Power Consumption Analysis

Table IV shows the power consumption of our prototype. The table shows how the power consumption of the whole system in *sleep* mode is more than 10 times less than that in *normal* mode. We estimated that the average time to elaborate one abandoned/removed object event is about 3 s from the moment when the detection starts. Moreover, the amount of time required for blob analysis depends on the number and size of the detected ROIs. Hence, it will be zero if the system does not detect any blob and about 100 ms for three ROIs sized 16 × 16 pixel. Table V shows the time request from the microcontroller to carry out the entire

TABLE IV
POWER CONSUMPTION OF THE VIDEO SENSOR NODE

Component	Power [mW]
ARM9 (RUN mode)	450
ARM9 (IDLE mode)	49.5
ARM9 (SLEEP mode)	15
Video sensor (ON mode)	165
Video sensor (IDLE mode)	23
TX/RX (ACTIVE mode)	98
TX/RX (IDLE mode)	10
PIR sensor	1.5
VIDEO NODE	
Active with video sensor	626.5
Active, without video sensor	484.5
Alarms Transmission	572.5
Sleep, only PIR is Active	51

TABLE V
ENERGY AND TIME REQUIREMENTS OF THE PERFORMED TASKS

Task	Energy [mJ]	Time
Frame Acquisition	58.5	93.5
NCC Background Subtraction	455.8	940
Labeling	29	30
Blob Analysis	0 - 48,6	0-95
Image Transmission	601,1	1050

battery during the day, the energy accumulated in the battery being sufficient to keep the node on during the night.

TABLE VI
BATTERY LIFE OF THE VIDEO SENSOR NODE (IN HOURS)

No PIR	50 ev./h	100 EV./H	300 EV./H
13h	75h	48h	20h

video processing task. To evaluate the benefits provided by our approach in terms of improvement of battery life we consider the same system without PIR sensor. More specifically, we compare the two approaches assuming different amounts of random events (i.e., 50, 100, 300 events per hour) and by estimating the lifetime of the three scenarios using a full 2000 mAh battery. In the first simulation the PIR sensor is disconnected and the system does not know when to switch off/on the camera or when to turn into sleep mode, so it will work continuously and the power consumption will be independent from the number of events. Then, in our approach, and as previously mentioned, the PIR sensor detects events and wakes up the microcontroller which waits until the FOV is empty to start the video analysis. When the video analysis is over the microcontroller turns the camera off to reduce power consumption. For these simulations we assume that the amount of time that the microcontroller has to wait before starting the video analysis is 5 s. To evaluate this time we collected observations concerning the number of people entering our building through the main entrance, and we built a profile over 10 consecutive days. Successively, we computed the average time needed for the FOV to be empty again. The results comparing the life time (in hours) of both systems are shown in Table VI. From the table it is clear that the proposed approach allows for a substantial increase of battery life.

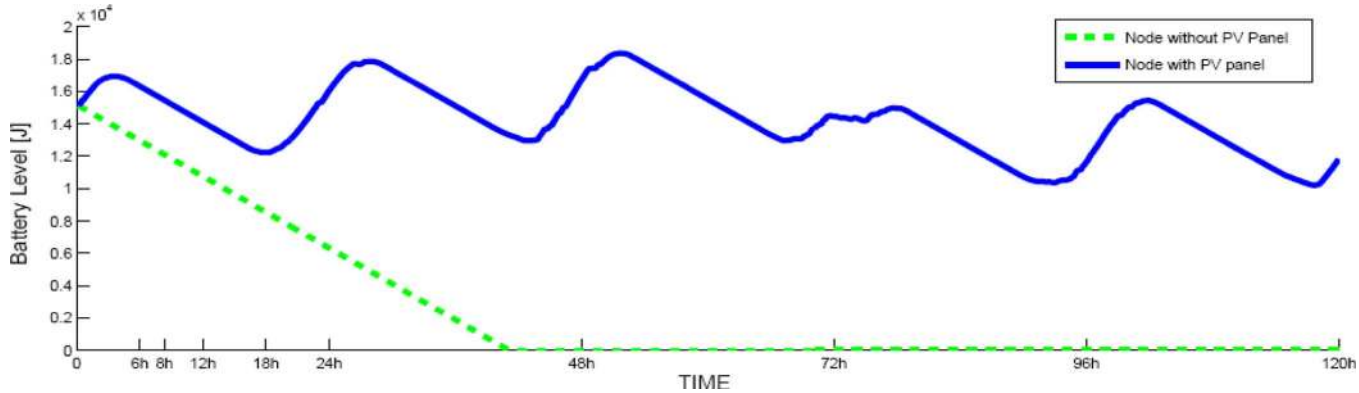

Fig. 11. Simulation of our system with and without PV harvester (50 events/h).

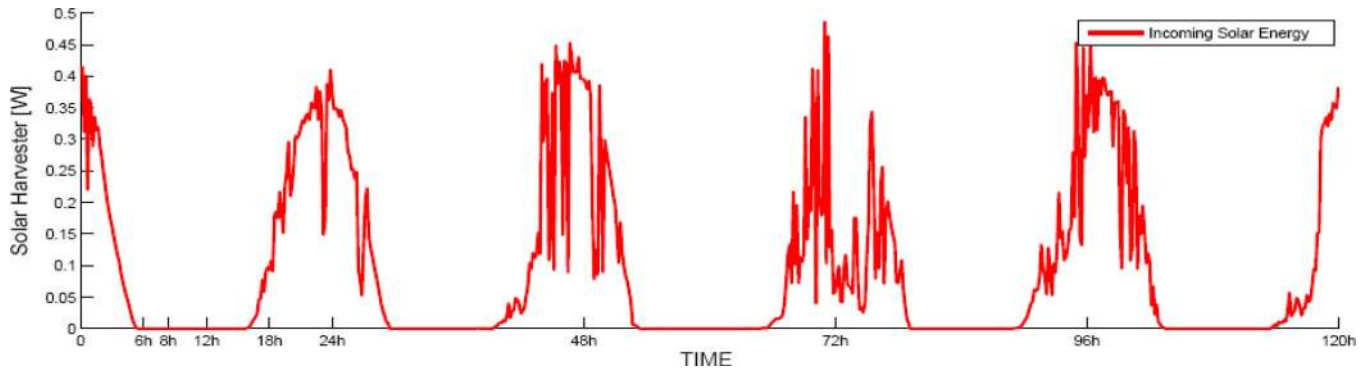

Fig. 12. Solar energy harvested during five consecutive days from our PV panel.

Using the solar panel to harvest the energy, we can obtain an additional improvement of battery life, achieving perpetual working in an application scenario with 50 events per hour. A simulation verifying the performance of the proposed system is depicted in Fig. 11, where the energy harnessed from the solar cell powers the sensor node and recharges the battery with the exceeding events. In our simulations we assume a rate of 50 events per hour representing people walking within the field of view of the video node. Furthermore we measured the energy intake from the energy harvester and the solar light intensity during five days (Fig. 12). All the information was stored in files used as input to our simulations. Finally the energy level of the battery at the beginning of the simulation is half its full capacity (2000 mAh, expressed in Joules). This simulation shows how the system can work perpetually with a small PV panel. The node without the energy harvester stops working after approximately 40 h, as it can be seen from Fig. 11. Instead, the node with the harvester can recharge its battery during the day, the energy accumulated in the battery being sufficient to keep the node on during the night.

VI. Conclusion

In this work, a self-powered wireless smart camera for run time video processing has been presented. The proposed approach have shown that combing advanced hardware/software solutions makes possible high accuracy and perpetual work in video processing on wireless sensor network where smart cameras find the most critical constrain in the power consumption. In fact, energy harvesters provide energy to replenish the batteries, while the combined use of different sensors with heterogeneous features allows for a remarkable reduction of the overall power consumption. We designed the camera for real time video processing algorithm and an ad-hoc abandoned/removed object using PIR features. Data were presented to evaluate the performance of our approach in term of energy consumption, video processing application and accuracy. The experimental results showed the versatility of the application in indoor and outdoor scenario and the simulation showed the perpetual work was achieved in simulation outdoor scenario. Future work will include a complete campaign of new experiments in order to analyze better the issue of background changes and more challenging condition (i.e., direct sun light, shadows). Moreover they will be directed towards improving the hardware of the smart camera node, video processing algorithm, and distributed intelligence.

References

[1] W. Chan, J. Chang, T. Chen, Y. Tseng, and S. Chien, "Efficient content analysis engine for visual surveillance network," *Trans. Circuits Syst. Video Technol.*, vol. 19, no. 5, pp. 693–703, 2009.

[2] M. Magno, D. Brunelli, P. Zappi, and L. Benini, "A solar-powered video sensor node for energy efficient multimodal surveillance," in *Euromicro Conf. Digital System Design*, 2008, pp. 512–519.

[3] A. S. Weddell, M. Magno, G. V. Merrett, D. Brunelli, B. M. Al-Hashimi, and L. Benini, "A survey of multi-source energy harvesting systems," in *Design, Automat. Test Eur. Conf. Exhibit.*, Mar. 18–22, 2013, pp. 905–910.

[4] S. Lu, J. Zhang, and D. Feng, "A knowledge-based approach for detecting unattended packages in surveillance video," in *Proc. IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS 06)*, Nov. 2006, p. 110.

[5] T. Semertzidis, K. Dimitropoulos, A. Koutsia, and N. Grammalidis, "Video sensor network for real-time traffic monitoring and surveillance," *IET Intell. Transport Syst.*, vol. 4, no. 2, pp. 103–112, Jun. 2010.

[6] S. Lim and L. Davis, A one-threshold algorithm for detecting abandoned packages under severe occlusions using a single camera CS Dept., Univ. Maryland, Tech. Rep. CS-TR-4784, 2006.

[7] S. Ferrando, G. Gera, and C. Regazzoni, "Classification of unattended and stolen objects in video-surveillance system," in *Proc. IEEE Int. Conf. Adv. Video Signal Based Surveill.*, Nov. 2006, p. 21.

[8] J. San Miguel and J. Martnez, "Robust unattended and stolen object detection by fusing simple algorithms," in *Proc. IEEE Int. Conf. Adv. Video Signal Based Surveill.*, Sep. 2008, pp. 18–25.

[9] M. Bhargava, C. Chen, M. Ryoo, and J. Aggarwal, "Detection of abandoned objects in crowded environments," in *Proc. IEEE Conf. Adv. Video Signal Based Surveill.*, Sep. 2007, pp. 271–276.

[10] F. Porikli, Y. Ivanov, and T. Haga, "Robust abandoned object detection using dual foregrounds," *EURASIP J. Adv. Signal Process.*, vol. 2008, no. 1, Jan. 2008.

[11] M. Spengler and B. Schiele, "Automatic detection and tracking of abandoned objects," in *Proc. IEEE Int. Workshop Vis. Surveill. Performance Eval. Track. Surveill.*, 2003.

[12] Y. E. A. Tian, "Robust and efficient foreground analysis for real-time video surveillance," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2005, pp. 1182–1187.

[13] P. Spagnolo, A. Caroppo, M. Leo, T. Martiffigiano, and T. D'orazio, "An abandoned/removed objects detection algorithm and its evaluation on PETs datasets," in *Proc. IEEE Conf. Adv. Video Signal Based Surveill.*, Nov. 2006, p. 17.

[14] M. Beynon, D. Van Hook, M. Seibert, A. Peacock, and D. Dudgeon, "Detecting abandoned packages in a multi-camera video surveillance system," in *Proc. IEEE Conf. Adv. Video Signal Based Surveill.*, Jul. 2003, pp. 221–228.

[15] C. Sacchi and C. Regazzoni, "A distributed surveillance system for detection of abandoned objects in unmanned railway environments," *IEEE Trans. Veh. Technol.*, vol. 49, no. 5, pp. 2013–2026, 2000.

[16] N. Bird, S. Atev, N. Caramelli, R. Martin, O. Masoud, and N. Papanikolopoulos, "Real time, online detection of abandoned objects in public areas," in *Proc. IEEE Conf. Robot. Automat.*, 2006, pp. 3775–3780.

[17] M. Magno, F. Tombari, D. Brunelli, L. Di Stefano, and L. Benini, "Multi-modal video surveillance aided by pyroelectric infrared sensors," in *Proc. ECCV Workshop Multi-Camera Multi-Modal Sensor Fusion, Algor. Appl.*, 2008, pp. 1–12.

[18] D. Carli, D. Brunelli, D. D. Bertozzi, and L. Benini, "A high-efficiency wind-flow energy harvester using micro turbine," in *SPEEDAM 2010—Int. Symp. Power Electron., Electr. Drives, Automat. Motion*, 2010, pp. 778–783.

[19] C. Moser, D. Brunelli, L. Thiele, and L. Benini, "Real-time scheduling with regenerative energy," in *Euromicro Conf. Real-Time Syst.*, 2006, pp. 261–270.

[20] C. Moser, L. Thiele, D. Brunelli, and L. Benini, "Robust and low complexity rate control for solar powered sensors," in *Proceedings -Design, Automation and Test in Europe, DATE*, 2008, pp. 230–235.

[21] P. de la Hamette *et al.*, "Architecture and applications of the fingermouse: A smart stereo camera for wearable computing HCI," *Personal Ubiquitous Comput.*, vol. 12, no. 2, pp. 97–110, 2008.

[22] D. Corley and E. Jovanov, "A low power intelligent video-processing sensor," in *Proc. 34th Southeastern Symp. System Theory*, 2002, pp. 176–178.

[23] D. Li, Y. Jiang, and G. Chen, "A low cost embedded color vision system based on SX52," in *Proc. IEEE Int. Conf. Inf. Acquisit.*, Aug. 2006, pp. 883–887.

[24] G. Zhang, T. Yang, S. Gregori, J. Liu, and F. Maloberti, "Ultra-low power motion-triggered image sensor for distributed wireless sensor network," in *Proc. IEEE Sensors*, Oct. 2003, vol. 2, pp. 1141–1146.

[25] W.-C. Feng, B. Code, E. Kaiser, M. Shea, W.-C. Feng, and L. Bavoil, "Panoptes: Scalable low-power video sensor networking technologies," in *MULTIMEDIA'03: Proc. 11th ACM Int. Conf. Multimedia*, New York, 2003, pp. 562–571.

[26] L. Ferrigno and A. Pietrosanto, "A low cost visual sensor node for bluetooth based measurement networks," in *Proc. 21st IEEE Instrum. Meas. Technol. Conf.*, 2004, vol. 2, pp. 895–900.

[27] P. Garda, O. Romain, B. Granado, A. Pinna, D. Faura, and K. Hachicha, "Architecture of an intelligent beacon for wireless sensor networks," in *Proc. IEEE 13th Workshop Neural Netw. Signal Process.*, 2003, pp. 151–158.

[28] D. C. N. Drajic, "Adaptive fusion of multimodal surveillance image sequences in visual sensor networks," *IEEE Trans. Consumer Electron.*, vol. 53, no. 4, pp. 1456–1462, Nov. 2007.

[29] Y. S. K and Y. Ho, "An efficient image rectification method for parallel multi-camera arrangement," *IEEE Trans. Consumer Electron.*, vol. 3, no. 57, pp. 1041–1048, Aug. 2011.

[30] T. He, P. Vicaire, T. Yan, Q. Cao, C. Zhou, L. Gu, L. Luo, R. Stoleru, J. A. Stankovic, and T. F. Abdelzaher, "Vigilnet: An integrated sensor network system for energy efficient surveillance," *ACM Trans. Sen. Netw.*, vol. 2, no. 1, pp. 1–38, Feb. 2006.

[31] A. Rajgarhia, F. Stann, and J. Heidemann, "Privacy-sensitive monitoring with a mix of IR sensors and cameras," in *Proc. 2nd Int. Workshop Sensor Actor Netw. Protocols Appl.*, Aug. 2004, pp. 21–29.

[32] S. Ren, Q. Li, H. Wang, and X. Zhang, "Design and analysis of wave sensing scheduling protocols for object-tracking applications," in *DCOSS Distributed Computing of Sensor Systems*. New York: Springer, 2005, pp. 228–243.

[33] C. Gui and P. Mohapatra, "Virtual patrol: A new power conservation design for surveillance using sensor networks," in *Proc. 4th Int. Symp. Inf. Process. Sensor Netw.*, 2005, pp. 246–253.

[34] C. Tsai, Y. Bai, C. Chu, C. Chung, and M. Lin, "Pir-sensor-based lighting device with ultralow standby power consumption," *IEEE Trans. Consumer Electron.*, vol. 3, no. 57, pp. 1157–1164, Aug. 2011.

[35] Y. Bai, Z. Xie, and Z. Li, "Design and implementation of a home embedded surveillance system with ultra-low alert power," *IEEE Trans. Consum. Electron.*, vol. 57, no. 1, pp. 153–159, Feb. 2011.

[36] D. Niyato, E. Hossain, and A. Fallahi, "Sleep and wakeup strategies in solar-powered wireless sensor/mesh networks: Performance analysis and optimization," *IEEE Trans. Mobile Comput.*, vol. 6, no. 2, pp. 221–236, Feb. 2007.

[37] P. Muralt, Micromachined infrared detectors based on pyroelectric thin films P Swiss Fed Inst Technol, EPFL, Dept. Mat. Sci., Ceram. Lab., Swiss Fed, Inst. Technol., EPFL, Dept. Mat. Sci., Ceram. Lab.,, Rep. Prog. Phys., 2001, vol. 64, pp. 1339–1388.

[38] R. Whatmore, "Pyroelectric devices and materials," *Rep. Progr. Phys.*, vol. 49, no. 12, pp. 1335–1386, 1986.

[39] P. Elmer, "Frequency range for pyroelectric detectors," [Online]. Available: http://www.perkinelmer.com

[40] D. Brunelli, D. Dondi, A. Bertacchini, L. Larcher, P. Pavan, and L. Benini, "Photovoltaic scavenging systems: Modeling and optimization," *Microelectron. J.*, vol. 40, no. 9, pp. 1337–1344, Sep. 2009.

[41] D. Dondi, A. Bertacchini, Larcher, P. Pavan, D. Brunelli, and L. Benini, "A solar energy harvesting circuit for low power applications," in *Proc. IEEE Int. Conf. Sustainable Energy Technol.*, Nov. 24–27, 2008, pp. 945–949.

[42] D. Porcarelli, D. Brunelli, M. Magno, and L. Benini, "A multi-harvester architecture with hybrid storage devices and smart capabilities for low power systems," in *Proc. 21st Int. Symp. Power Electron., Electr. Drives, Automat. Motion*, 2012, pp. 946–951.

[43] S. Huang, "An advanced motion detection algorithm with video quality analysis for video surveillance systems," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, pp. 1–14, 1, Jan. 2011.

[44] W. Wang, J. Yang, and W. Gao, "Modeling background and segmenting moving objects from compressed video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 5, pp. 670–681, May 2008.

[45] J. Suhr, H. Jung, G. Li, and J. Kim, "Mixture of Gaussians-based background subtraction for Bayer-pattern image sequences," *Trans. Circuits Syst. Video Technol.*, vol. 21, no. 3, pp. 365–370, Mar. 2011.

[46] F. Tombari, L. Di Stefano, and S. Mattoccia, "A robust measure for visual correspondence," in *Proc. Int. Conf. Image Anal. Process.*, 2007, pp. 376–381.

[47] R. Zabih and J. Woodfill, "Non-parametric local transforms for computing visual correspondence," in *Proc. Eur. Conf. Comput. Vis.*, 1994, pp. 151–158.

[48] D. Bhat and S. Nayar, "Ordinal measures for image correspondence," *IEEE Trans. Pattern Recognit. Mach. Intell.*, vol. 20, no. 4, pp. 415–423, Apr. 1998.

[49] A. Mittal and V. Ramesh, "An intensity-augmented ordinal measure for visual correspondence," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, 2006, pp. 849–856.

[50] B. Xie, V. Ramesh, and T. Boult, "Sudden illumination change detection using order consistency," *Image Vis. Comput.*, vol. 2, no. 2, pp. 117–125, 2004.

[51] N. Ohta, "A statistical approach to background subtraction for surveillance systems," in *Proc. Int. Conf. Comput. Vis.*, 2001, vol. 2, pp. 481–486.

[52] B. Xie, V. Ramesh, and T. Boult, "Sudden illumination change detection using order consistency," *Image Vis. Comput.*, vol. 2, no. 2, pp. 117–125, 2004.
[53] A. Lanza and L. Di Stefano, "Detecting changes in grey level sequences by ml isotonic regression," in *Proc. AVSS*, 2006, p. 4.
[54] L. Di Stefano and A. Bulgarelli, "A simple and efficient connected components labeling algorithm," in *Proc. Int. Conf. Image Anal. Process.*, 1999, pp. 322–327.

Michele Magno (M'13) is a Postdoctoral Fellow at the MicrelLab Group, University of Bologna, Bologna, Italy, where he received the M.S. degree and the Ph.D. degree in electronic engineering.

The most important themes of his research are on power management techniques and extension of the lifetime of wireless sensor networks. In this field, he has worked actively both on the energy efficiency of the nodes and the network and on the use of harvesters, such as solar panels and wind, and using the fuel cells to hydrogen to feed the nodes and recharge the batteries. He has collaborated with several universities and research center such as ETH Zurich, University of Cork and Tyndall Institute, University of Trento, Politecnico di Torino. He has published 20 papers in international journals and conferences.

Federico Tombari (M'09) holds an appointment as a Postdoctoral Fellow at the Computer Vision Lab, University of Bologna, Bologna, Italy, where he received the B.Eng., M.Eng., and Ph.D. degrees in 2003, 2005, and 2009, respectively.

His current research activity concerns computer vision and robotic perception, and it encompasses co-authoring more than 50 refereed papers on peer-reviewed international conferences and journals, mainly focused on 2D/3D object recognition, stereo vision, video analysis for surveillance and efficient indexing. Since 2012 he also serves as an Adjunct Professor at the University of Bologna. Since 2012 he is a Senior Scientist Volunteer at Open Perception foundation, where he contributes to development, dissemination, and mentoring for the Point Cloud Library.

Dr. Tombari was the recipient of the "Best Paper Award Runner Up" of the international conference 3DIMPVT 2011. He is member of IAPR-GIRPR.

Davide Brunelli (M'10) received the M.S. (*summa cum laude*) and Ph.D. degrees from the University of Bologna, Bologna, Italy, in 2002 and 2007, respectively, both in electrical engineering.

He is Assistant Professor at University of Trento, Trento, Italy, since 2010. From 2005 to 2007 he was an academic guest at the Swiss Federal Institute of Technology (ETH), Zurich, Switzerland. He was leading industrial cooperation activities with Telecom Italia. His research interests concern the development of new techniques of energy scavenging for WSNs and embedded systems, optimization of low-power and low-cost WSNs, interaction and design issues in embedded personal and wearable devices.

Luigi Di Stefano (M'08) received the degree in electronic engineering, in 1989, from the University of Bologna, Bologna, Italy, where he received the Ph.D. degree in electronic engineering and computer science from the Department of Electronics, Computer Science and Systems, in 1994.

In 1995, he was postdoctoral fellow at Trinity College Dublin. He is currently Associate Professor at the Department of Computer Science and Engineering, University of Bologna, Bologna, Italy. His research interests include computer vision, image processing, and computer architecture. He is author of more than 130 papers and of several patents. Since 2012, he has been a member of the Scientific Advisory Board for the Datalogic Group.

Dr. Stefano is a member of the IAPR-IC.

Luca Benini (F'07) received the Ph.D. degree from Stanford University, Stanford, CA, USA, in 1997.

He is Full Professor of Digital Integrated Circuits and Systems at ETHZ D-ITET, and a Professor of Electronics at the University of Bologna, Bologna, Italy. He is also currently serving as Chief Architect in STMictroelectronics (AST division, Grenoble). His research interests are in the design of SoC platforms for embedded applications. He published more than 600 papers in peer-reviewed International Journals and Conferences, four books and several book chapters (h-index = 68 on Google Scholar). He is Associate Editor or the ACM *Transactions on Embedded Computing Systems*.

Dr. Benini is Associate Editor of the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS. He is an ERC Advanced Grant winner. He is a member of the Academia Europaea, a past member of the steering board of ARTEMISIA.