Using a Wireless Sensor Network to Enhance Video Surveillance

Rita Cucchiara, Andrea Prati, Roberto Vezzani, Luca Benini, Elisabetta Farella, Piero Zappi

***Abstract*— To enhance video surveillance systems, multi-modal sensor integration can be a successful strategy. In this work, a computer vision system able to detect and track people from multiple cameras is integrated with a wireless sensor network mounting PIR (Passive InfraRed) sensors. The two subsystems are briefly described and possible cases in which computer vision algorithms are likely to fail are discussed. Then, simple but reliable outputs from the PIR sensor nodes are exploited to improve the accuracy of the vision system. In particular, two case studies are reported: the first uses the presence detection of PIR sensors to disambiguate between an open door and a moving person, while the second handles motion direction changes during occlusions. Preliminary results are reported and demonstrate the usefulness of the integration of the two subsystems.**

***Index Terms*— Sensor Network, PIR, multiple cameras, tracking, video surveillance**

I. Introduction

Video surveillance and other security-related applications have gained many credits due to the terroristic threats of the last years. Several industrial and academic projects have recently started to increase the accuracy of (semi-)automatic surveillance systems. In addition, the abatement of hardware costs allows the deployment of thousands of cameras for surveillance purposes at a reasonable cost.

The ever-increasing demand of security and the low cost of cameras contributed to the diffusion of the research in *distributed multi-camera surveillance systems*. Multiple cameras enable the surveillance of wider areas and the exploitation of redundant information (provided by the different viewpoints) might solve classical limitations of single-camera systems, such as occlusions.

Despite the efforts made by the researchers in developing a robust multi-camera vision system, computer vision algorithms have proved their limits to work in complex and cluttered environments [1]. These limits are mainly due to two classes of problems. The first is that “non-visible areas can not be processed by the system”. This trivial statement is of particular importance in cluttered scenes and can be partially lessened by using multiple sensors (not only cameras). The second class of problems, instead, is due to the limited resolution of cameras. Having infinite resolution and zooming capabilities would make the job easier, but, in addition to be unfeasible, it would exponentially increase the computational load and it is typically too expensive.

Rita Cucchiara, Andrea Prati, and Roberto Vezzani are with University of Modena and Reggio Emilia, Italy. Luca Benini, Elisabetta Farella, and Piero Zappi are with University of Bologna, Italy. Andrea Prati is the corresponding author.

An interesting solution is that of using simple but effective specialized sensors to solve the specific problems of the vision systems. In this way, vision would still provide high-level information, and low-level sensors would assure higher accuracy. In this context, the marriage between a widely distributed low-cost wireless sensor network and the coarsely distributed higher level of intelligence that can be exploited by computer vision systems may overcome many troubles in a complete tracking of large areas. For our application, we exploit Passive Infrared (PIR) sensors which are widely deployed in low-cost surveillance systems (e.g., for car or home alarm systems). PIR sensors are used in traditional surveillance systems to trigger video cameras [2]. A trigger just conveys a binary (yes/no) presence information, but limited signal processing effort on the output of a PIR sensor can produce much more information (e.g., target speed and direction of movement). Furthermore, integration of data from multiple networked PIR sensors can provide drastically improved spatial resolution in monitoring and tracking. Low-cost and low-power sensor nodes can now be developed with on-board processing capabilities, reconfigurability and wireless connectivity [3], [4].

This work reports our joint research in developing a multimodal sensor network that integrates a wireless network of PIR-based sensors with a traditional vision system to provide drastically improved (in accuracy and robustness) tracking of people. It is worth noting that people surveillance is more interesting from the research's point of view w.r.t. to vehicle tracking, because of the intrinsic complexity in detecting, tracking, and understanding human behaviour: changes in posture and gestures, human interaction, presence of multiple people, and so on, make the problem challenging and interesting for the computer vision community.

The paper is structured as follows. The next section analyses the related works on the topics of distributed video surveillance, multi-modal sensor integration and PIR sensors presented in the literature. Section III reports the software architecture of the system and the integration of the PIR sensor network and the computer vision subsystems, described, respectively, in sections IV and V. The preliminary results are provided in section VI, by means of two case studies. Eventually, conclusions are drawn.

II. Related works

The emerging technology of multisensor data fusion has a wide range of applications, both in public and in private areas. Among them, surveillance and security have gained much interest in recent years, also due to the terroristic threats. Such applications often rely on cameras, due to the large amount of information and the high-level knowledge they provide. However, video surveillance in wide area through computer vision techniques is still a challenging problem largely faced in the last years [5]. In fact, additionally to the problems of segmentation and tracking from a single camera, it also requires a consistent labelling phase. In other words, the identities of the objects detected from each camera must be preserved in order to analyse their behaviours.

Adjacent cameras can have overlapped field of views or not. In the case of non-overlapped cameras the only feature usable for maintaining the identity of a moving subject is its visual appearance. To this aim, colour and shape ratios are widely adopted [6]. Unfortunately, using merely the subject's appearance is not a successful strategy, since the appearance (in particular, the colour) can be reproduced very differently with different cameras and under different illumination conditions.

Some works that use only subject's appearance have been proposed also in the case of overlapped cameras [7], [8]. However, in the case of partially overlapped cameras, the best choice is to exploit also geometrical information. In [9], each camera processes the scene and obtains a set of tracks. Then, regions along the epipolar lines in pairs of cameras are matched and the mid-points of the matched segments are back-projected in 3D and then, with an homography, onto the ground plane to identify possible positions of the person within a probability distribution map (filtered with a Gaussian kernel). A particularly interesting paper is reported in [10] in which homography is exploited to solve occlusions. Single camera processing is based on particle filter and on probabilistic tracking based on appearance to detect occlusions.

Most of these approaches require camera calibration. In outdoor environments with many cameras, placed in high positions over poles at unknown distance, manual calibration could be difficult and time consuming to achieve. Thus, automatic camera calibration techniques have been proposed. Relevant examples of these are the work of Jaynes [11] and Khan and Shah [12]. The first one supposes that the trajectory of the people during the transition between a camera's field of view to that of another camera keeps the same direction and speed. With this assumption, the paper shows a method to register the two views. The approach of Khan and Shah, instead, is based on the computation of the so-called *Edges of Field of View* (EoFoV, hereinafter), i.e. that are the projections on one camera of the limits of the field of view of the other camera, and, thus, defining the overlapped regions. Our algorithm for consistent labelling follows the same geometrical approach of this proposal to compute, starting from the EoFoV lines extraction, the homographic relationship between the two ground planes in an automatic way. This transformation is then exploited for establishing the consistent labelling.

Since cameras still have limitations due to the limited field of view or resolution, the proposal of this work is to combine data coming from cameras with information provided by a wireless sensor network based on Passive Infrared Sensors (PIR), in order to improve the robustness of the system. PIR sensors are mainly known for their use in video surveillance systems, manufactured by a number of companies (e.g in [2], [13]), to detect motion and provide a trigger to cameras. For their insensitivity to light conditions, in [14] PIR sensors are used to provide a trigger event in a motion-detection application mainly based on cameras for tracking events at night, together with a floodlight. The appearance of an infrared radiating body set off the PIR sensor, which turns on the floodlight enabling the cameras to capture clearly an event such as animals passing by an outdoor detected area. Being low-cost, low-power and small form-factor devices, PIR sensors are suitable for wireless sensor networks, where energy consumption, unobtrusiveness and easy placement in the environment are critical requirements. In [15], a wireless PIR sensor network is used to detect objects and humans for security applications and provide an estimation of the direction of movements. The network is implemented using Mica2 [4] nodes and data gathered by a base station. Tracking algorithms are implemented on the nodes and speed calculation provided accurately, even if influenced by the orientation of the sensors. Sensor networks implemented with PIR are useful where privacy must be preserved together with security. In [16] cameras and PIR sensors are deployed respectively in public and private areas, and their information combined to correlate events such as tracking human motion and undesired access or presence in private areas, such as theft. This work demonstrate benefits of reducing camera deployment in favor of PIR sensors and reports results from a survey on 60 people, stating that people consider motion sensors less invasive for their privacy than cameras.

PIR sensors are often combined with vision systems and other kind of sensors in research focused on robot navigation and localisation. In [17] a sound source localiser and a motion detector system are implemented on a human service robot called ISAC, with the purpose of redirect the attention of ISAC. The motion detector system use an infrared sensor array of five PIR sensors and it is integrated with the vision system of ISAC to perform real-time human tracking, in a most inexpensive way. Similar use of PIR sensor can be found in [18], where data from multiple PIR sensors, two colour cameras and a pair of microwave sensors are collected, processed and fused to track human presence in the vicinity of a robot. A main motivation to use PIR sensors in pair with cameras is their insensitivity to lighting conditions, to avoid robot collision with humans for their safety in many different conditions.

III. Integrated Multi-Modal Sensor Network

Vision systems achieve good accuracy when working alone, but they definitely could benefit from the multi-modal integration with PIR sensors. For testing the integration, a test bed has been created at our campus. Fig. 1 shows the location of cameras and PIR sensors. The system we implemented is composed by several modules, working in parallel on different threads (see Fig. 2). In particular, a thread is generated for each camera, devoted to compute the list of people present in the scene exploiting a two stage processing (segmentation and tracking). All the camera threads are tightly connected to a

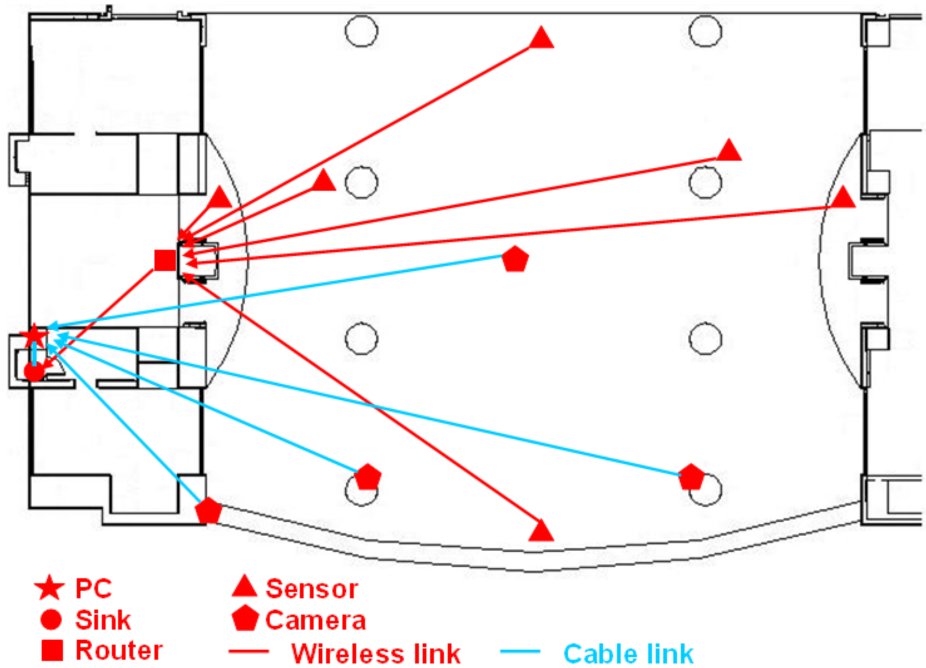


Fig. 1. Map of our test bed system

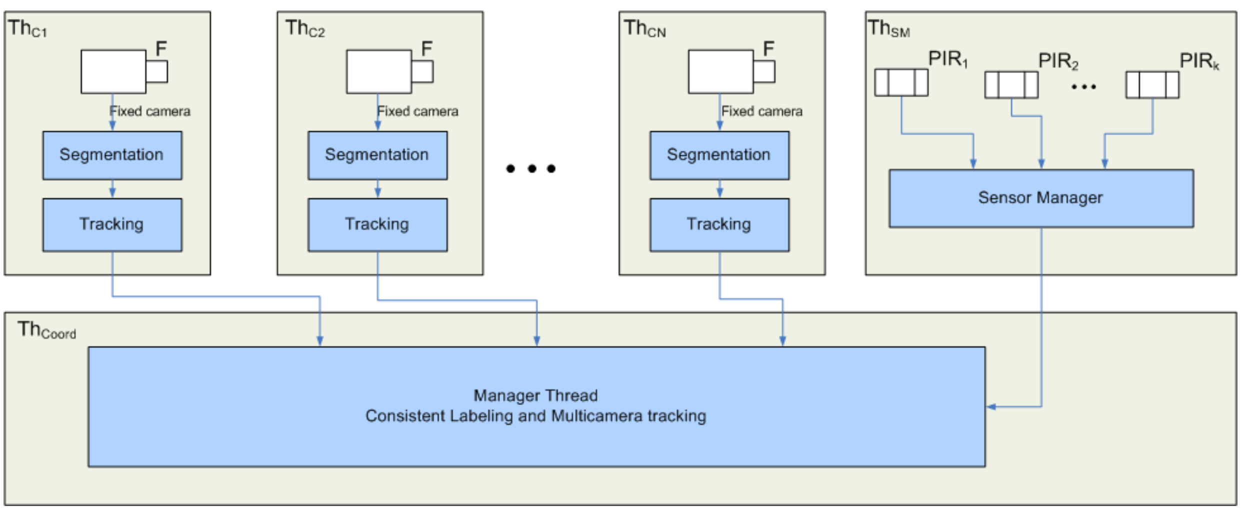


Fig. 2. Software architecture of the system

coordinator thread, that detects if the same person is visible in more than one camera, and, in such a situation, it labels the corresponding tracks with the same identifier.

At the same time, a sensor manager coordinates the network of sensors distributed over the monitored area. As explained in detail in Sec. IV, PIR sensors are able to detect the presence of moving objects or people within their coverage area. Observing the output of a couple of PIR, the microcontroller integrated on the sensor node is able to detect both presence and direction of movement of the person walking by it. When such situation is detected the microcontroller creates and wirelessly sends a message to a special node which acts as a sink. The sink then forwards the message to the sensor manager via RS232 cable in order to make the information available to the tracking and labelling algorithms.

Eventually, data coming from cameras and sensors are collected and managed by a supervisor thread. The coordination between cameras and sensors is twofold. Each time the vision system requires more detailed or reliable information about the presence of people in the zones monitored by the sensors, it sends requests to the supervisor thread. Contemporarily, when the sensor network detects a particular event, the manager takes care to inform the involved cameras. Detailed descriptions of the PIR sensor network, of the vision system, and of two examples of integration are reported in the following sections.

IV. PIR SENSOR NETWORK

A. Sensor node architecture

Sensor nodes are more complex devices than simple sensors. A microcontroller, a transceiver and a power supply (mainly a battery) together with a variety of sensors form an entity capable of collecting events and information from the surrounding area, analyse them and send message or data to other nodes or to the user (see Fig. 3).

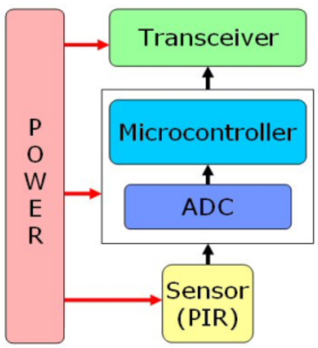


Fig. 3. Sensor node architecture

Microprocessor and Transceiver. The hardware used to collect, elaborate and send data from the PIR sensor is the SARD™Board provided by Freescale™to develop wireless applications over the worldwide free 2.4 GHz ISM band. This development board includes a microcontroller of the HCS08 family (MG9S08GT60), an MC13192 2.4GHz transceiver, a set of I/O pins and a RS232 port to interface the CPU with the external world and several leds and buttons. This development board is very flexible and many types of sensors can be connected to it. Furthermore, the microcontroller has the ability to operate in low power mode to save energy and to be waken up by external interrupts generated by the sensor output conditioning circuits. The transceiver is designed to be low-power, short range and fully compatible with the IEEE 802.15.4 standard. It communicates with the microcontroller via a four-wire SPI interface and an interrupt request output. The transceiver supports 250 kbps O-QPSK data in 16 5.0MHz channels and full spread-spectrum encode and decode. It enables communication with other devices up to distances of 40m outdoor.

Sensing element. The sensing element used in our application is a passive infrared (PIR) sensor. Such devices are able to transduce incident infrared radiation into current. Commercial PIRs are sold in pairs with opposite polarization. Such configuration makes the sensor able to detect variations of incident infrared radiation irradiated by bodies moving inside his coverage area which are not at thermal equilibrium with the environment. PIR are used with Fresnel lens to enlarge and shape their area of sensitivity. They are passive sensors with minimal power consumption, ideal for battery powered systems. The PIR sensor used in this work exhibits high sensitivity and reliable performance, high stability to thermal change and high immunity to external noise. By suitably shading its Fresnel lenses we were able to obtain cone of coverage with a vertical angle of 60 degrees and an horizontal angle of 38 degrees.

Furthermore, a single PIR sensor can detect the direction of movement. Fig. 4 shows the signal detected by sensor when a person passes through the area under control from left to right (the first peak is negative) and from right to left (first peak is positive).

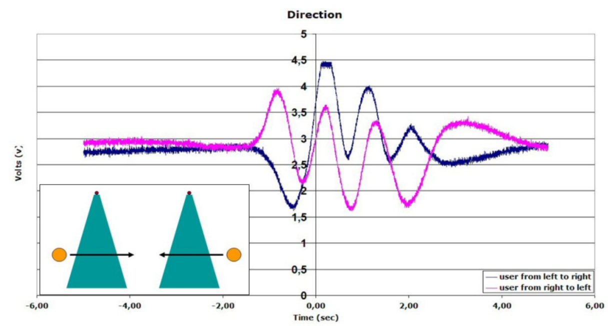


Fig. 4. Signals detected by sensor: when a person passes through the area under control from left to right the first peak is negative and from right to left the first peak is positive

In our setup, the typical node includes two PIR sensors to enhance the information captured as explained in the following paragraph.

Power Supply. The system can be powered directly by main power or alternatively with a commercial battery at 9V. Internally, the voltage is stabilised at 3.3V.

B. Sensing and Acquisition Software

Fig. 5 shows the processing data flow from event acquisition to generation of the packet which will be sent by the wireless node. This section describes the role of the different parts.

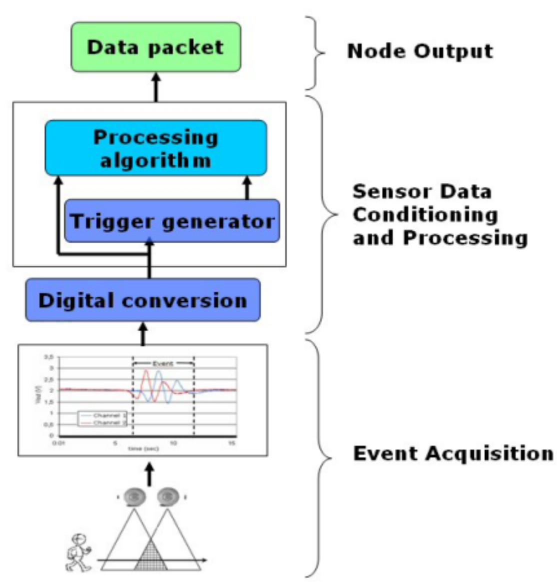
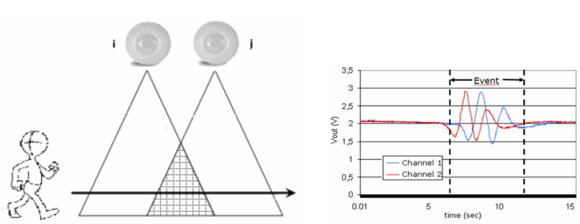


Fig. 5. Event acquisition and Sensor Data Conditioning and Processing

We are interested in detecting precisely presence and direction of movement, but also more complex movements such as changes in direction within the covered area. In fact, these are information that can be exploited by the vision system for enhancing the accuracy of the video surveillance application, in which presence and direction of movement (of people) are key information.

As outlined above, we augment the information produced by a single node by using 2 PIR sensors (Fig. 6(a)) per node. The typical sensors' output when a person is walking through the sensor area is the one presented in Fig. 6(b). The signal collected by the sensors is digitally converted to be processed by the microcontroller. When a person crosses the monitored area each of the two sensors generates a waveform similar to the one in Fig. 4 depending on the direction of movement. We consider interesting events those stimulating a significant variation of the signal (Fig. 6(b)): when the input coming from the digital converter exceeds a lower or an upper threshold, a trigger is generated to start the processing algorithm in charge of extracting information from the signal. The analysis, as mentioned above, is aimed at understanding the direction of a person walking in the covered area. Assuming that one person is moving from left to right as in Fig. 6(a), he will be detected first by PIR_i then by both PIR_i and PIR_j and at last only by PIR_j as it is lightened in Fig. 7. In general, a different activation sequence can help identifying changes in direction of movement within the area covered by the array of sensors. Results from the processing is a message containing information about the presence and/or direction of movements in the selected area. The format of the packet is described in the following section.



(a) (b)

Fig. 6. Sensor node composed by two PIRs

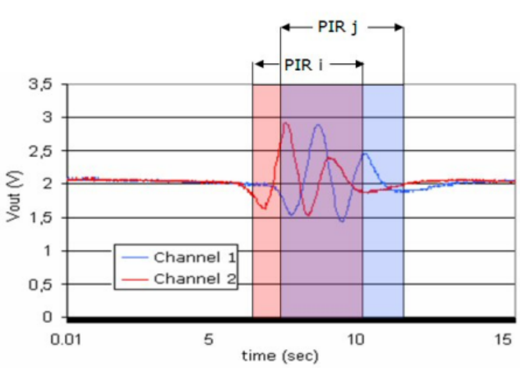


Fig. 7. Activation sequence.

Note that the trigger generator is disabled for a period to be set depending on the application after the detection of an event, avoiding redundant information to be sent. In our case, the period is set at 2 seconds. This choice has been verified as not influencing correct analysis, because it does not cause loss of events.

C. Network Architecture

Communication Protocol. Communication among nodes is based on IEEE 802.15.4 protocol for wireless networks. This protocol has been designed for applications with limited power budget and relaxed throughput requirements. Its main objectives are ease of installation, reliable data transfer, short range operation, extremely low cost and a reasonable battery life, while maintaining a simple and flexible protocol. This protocol defines two types of devices: full function device (FFD) and reduced function device (RFD). It also defines three possible roles for the nodes of the network: coordinator (which is unique for a network), router and end device. A coordinator and a router must be FFD while an end device may be either FFD or RFD. Usually routers and coordinators are main powered while end devices are located on the field and are battery powered. Two or more devices constitute a wireless personal area network (WPAN). However, one device must be a coordinator.

Network topology and organization. The network has a star topology, i.e., all the nodes are end devices and communicate only with a central one, the coordinator. The central node (bridge) collects data from the sensor nodes and sends them to another node (sink) which communicate through its RS232 interface to a PC (see Fig. 1). Hence, in our application the bridge is the network coordinator while the other nodes are end devices. The sensor nodes, which are located in the courtyard, are battery-powered while the bridge and the sink are main-powered. This topology is suitable to the characteristics of the monitored area. In fact, the sensor are located in a courtyard outside the building while the PC, due to privacy issues, is locked inside a small room, which must be kept closed within the building. Some tests shown that only the sensors close to the door of the building are able to communicate with a device inside the room, while all the courtyard can be covered by a receiver located close to the door.

Message format and set of commands. As already mentioned, the information collected by the sensors are sent to the video processing server via RS232 cable. We decided to use an asynchronous communication, that is, the sensor network send data to the server as soon as it collects them. The structure of the messages is shown in Fig. 8.

I	ID	AREA	LENGTH	[NAME	VALUE]$_n$	F

Fig. 8. Communication protocol between nodes and sensor manager

Each message is made up of a start byte (the ASCII code 'I'), a sensor ID, an area ID, an indication of length (the number of following couples name-value), several couples name-value and a stop bit (the ASCII code 'F'). Start and

Information	Code	Values	Code
Presence	1	Present	1
		Area free	16
Direction	2	From PIR_i to PIR_j	48
		From PIR_j to PIR_i	192

TABLE I

EXAMPLES OF ADOPTED CODES

stop bit are used for synchronization. Area and sensor node ID are used to uniquely identify the node. The couple name-value encodes the information provided by the sensor. Some examples are reported in Table I.

V. VISION SYSTEM

A. Single camera processing

Many approaches to people detection and tracking by single cameras have been proposed in the literature. Their schemes are often similar: first, to perform motion detection by separating points belonging to still parts from points belonging to moving parts (by means of background suppression, frame difference, or statistical analysis); then, blob analysis aims at grouping spatially correlated points into objects and characterizing them by visual features and motion components; eventually, moving objects are tracked with the aim of keeping track of their identity to further analyze the behaviour.

Our approach from single camera follows this scheme, and it is composed by two main modules: segmentation and tracking. The first module aims at extracting *visual objects*, that are entities that we are interesting in and that we want to analyze separately with respect to the background. Normally, the *visual objects* are objects detected in motion for almost one frame. To this aim, background suppression techniques are often adopted and operate by subtracting the current background model B^t from the current frame I^t. The points are extracted and grouped with a labelling process into a set FO^t of foreground objects at instant time t. This set contains both relevant objects and other outliers, such as shadows and noise. To identify shadow points, we used a deterministic approach, proposed in [19], based on the assumption that shadows have similar chromaticity but lower brightness than the background on which they are cast.

Objects in the set FO^t considered too small are discarded as noise. The set VO^t of visual objects obtained after the size-based validation is processed by the tracking module that computes for each frame t a set of tracks $T^t = \{T_1^t, ..., T_m^t\}$.

In the case of people tracking, the basic tracking approaches (based on directional rules, or Kalman filters) are not suitable, since humans undergo to deformation in the shape, move with unpredictability and sudden changes in the main direction, and are likely to be occluded by objects or other people. For these reasons, we proposed a probabilistic and appearance-based tracking algorithm able to manage also large and long-lasting occlusions [20]. Despite its accuracy, our tracking fails in the case the person changes his direction when occluded, since the algorithm relies on the hypothesis of constancy of motion during occlusions (being any other hypotheses not reasonable). Since in absence of visibility cameras are useless, this is a concrete and interesting example in which PIR sensors can be useful, as it will be shown in the following section.

The knowledge about VOs and their status is exploited by a selective background model [19] in order to be both reactive to background changes and robust to noise. Selective update is obtained by, on the one hand, not considering moving pixels in the updating process, and, on the other hand, forcing inclusion of stopped objects (previously moving) into the background model. Unfortunately, the system sometimes misclassifies moving objects (such as a person) with stopped objects (such as a door that has been opened). In these cases, the lack of enough resolution prevents the vision system to work properly and PIR sensors might help.

Eventually, scene understanding is a high-level module and heavily depends on the specific application. In the case of video surveillance of people, it includes a posture classification module [20], capable to discriminate between four postures (standing, sitting, crouching, and laying) and, consequently, to detect interesting events, such as a person's fall.

B. Consistent labelling

Real video surveillance setups often require multiple cameras, both to cover wider areas and to solve occlusions by exploiting multiple viewpoints. The goals of the consistent labelling and multicamera tracking module are the detection of correspondences between people extracted from each single camera tracking module, and then the computation of a list composed by the best views (selected from the different cameras) of people present in the scene. This list is the input of higher level tasks, as posture classification, face detection, and recognition. We propose an approach of consistent labelling based on geometrical features and homographic transformations. For two overlapped cameras C_i and C_j, through a learning procedure in which a single track moves from one view to another, an automatic procedure computes the EoFoVs that are exploited to keep consistent labels on the objects when they pass from one camera to the adjacent. By this, the homography that binds the ground planes on the two views can be easily computed. Full details can be found in [21].

Differently from other methods that check consistency only when objects pass through the edges of the field of views (camera handoff), we compute the assignment each time a new object is detected in the camera C_i in the overlapping. In this case its support point is projected in C_j by means of the homographic transformation. The coordinates of the projected point could not correspond to the support point of an actual object. Thus, we select for the match the object in C_j whose support point is at the minimum Euclidean distance in the 2D plane from these coordinates. This approach is an efficient trade-off between classical techniques that verify correspondences at the camera handoff instant only (as in [12]), and complex methods of 3D reconstruction that find correspondences at each frame preventing any real time implementation (as in [9]). Fig. 9 gives a bird-eye-view description of the area acquired by three different cameras; this representation is possible due to the homographic transformations between different views. The edges of field of view have been superimposed. The people can be detected by one, two, or even three cameras depending on their position. When a person is in the internal part (where three cameras are overlapped), three different views of the same person are available. In Fig. 10 an example of consistent labelling between three cameras is reported.

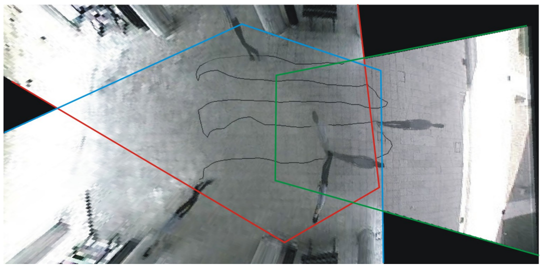


Fig. 9. Bird-eye-view description of our test bed environment

VI. Results of Multi-modal Integration

As stated above, the vision system achieves good accuracy when working alone, but it definitely could benefit from the multi-modal integration from PIR sensors. This section will report some preliminary results. To test the system we have equipped the atrium of our faculty with four cameras and several PIRs, as depicted in Fig. 1. Detailed descriptions with particular experimental results of the sensor nodes and of the multicamera system are reported in [22].

A. Sensor-guided Background Update

Fig. 11. Opening and closeing doors make unreliable background suppression techniques

Algorithms of motion capture based on background subtraction rely on a very crucial task: the update of the background, especially in presence of illumination changes and moved objects inside the scene. For example, when the doors in Fig. 11 are opened, the background scene changes and the detection of people in that area becomes unreliable. To this aim, we use sensors to monitor the area near the doors. If the single camera processing detects a visual object in the door area but the sensors do not capture events, then we assume that the motion is due to an incorrect background. In such a situation, the background is updated by forcing the area covered by the sensor directly with the input image.

More generally, each tracking system analyzes its list of detected objects. If an object is still for a long time, then the correspondent camera thread makes a request to the manager specifying the object location. The manager searches if the concerned zone is covered by a sensor and, in such a situation, it responds with the relative state. If the computer vision and the sensor network are discordant, then the sensor is considered more reliable, and the vision system reacts consequently, for example updating the background.

In Fig. 12 some frames taken from a single camera that is capturing the entrance of our faculty are reported. The rows report, from top to bottom, the input frames, the output of the tracking system, and the background model. Initially (first column) the door was open. Some frames later a person closes the door and from this instant the background becomes inconsistent. In fact, the system erroneously detects the presence of a person in the area of the door (see Fig. 12(e)). When the PIR sensor placed near the door does not capture any events, the background is correctly updated (last column).

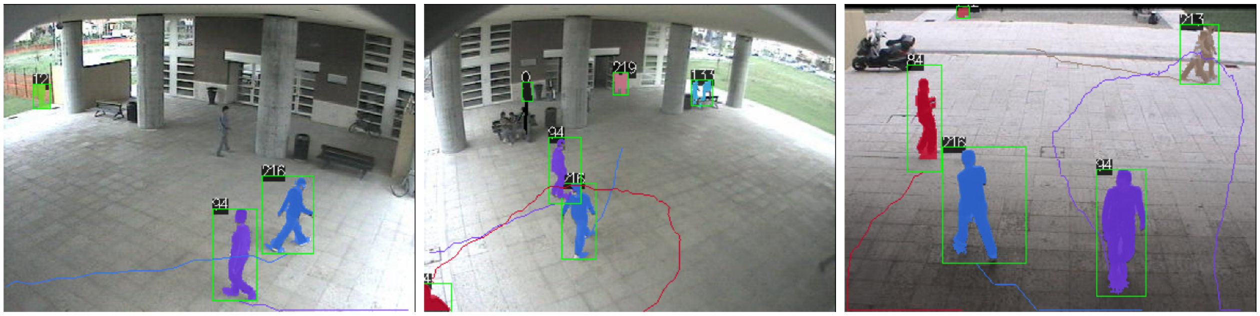
B. Detection of Direction Changes during Occlusion

Occlusions are another problem that characterises video-surveillance systems based on computer vision; for example, in the environments of Fig. 11, people can walk behind the columns, and, in such a situations, the system is likely to lose them. To face this problem, we have introduced some rules inside the tracking system. When a track disappears, it is not deleted immediately, but its appearance is kept unchanged and an estimation of the track position is computed exploiting a constant velocity assumption. If the person returns visible again with a similar appearance and a position near to the predicted one, then the system assigns the same label of the disappeared track. However, if the person changes direction during the occlusion, the system is not able to correctly assign the label anymore.

For this reason, we exploit a PIR sensor node placed behind the column. As above mentioned, these sensors detect not only the presence of a person, but also his direction. Then, we can detect a change of direction capturing couples of opposite direction events sent in a short temporal window. In such a situation, the direction of the motion applied to the track is inverted in order to estimate the position frame by frame.

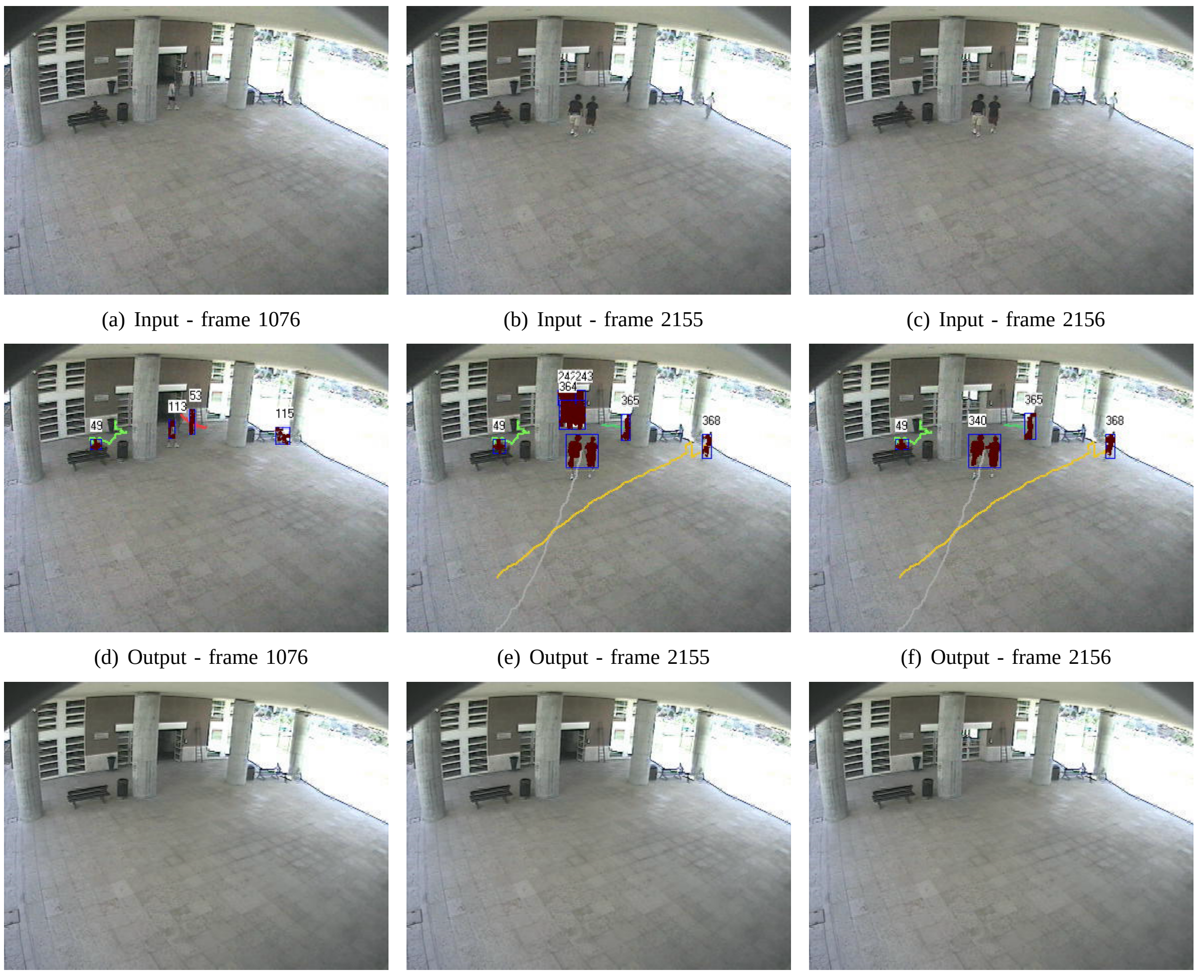
In Fig. 13 an example of consistent labelling after an occlusion is reported. The person walks behind a column and, during the occlusion, inverts his direction. The computer vision tracking algorithm is not able to solve the consistent labelling because the person reappears too far with respect to the predicted position (computed with a constant velocity assumption). Using PIR sensors, instead, the change of direction is detected and the estimated track position can be properly updated. Then, when the person reappears, the tracker assigns the same label (24) assigned before the occlusion (see Fig. 13(b)).

Differently from the previous example, in this case the sensor network detects an event and the manager thread informs the involved cameras of it. Then, if a tracking system has detected an object in the corresponding position, the motion direction is changed accordingly.



(a) C^1 (b) C^2 (c) C^3

Fig. 10. Example of consistent labelling between three views



(a) Input - frame 1076 (b) Input - frame 2155 (c) Input - frame 2156

(d) Output - frame 1076 (e) Output - frame 2155 (f) Output - frame 2156

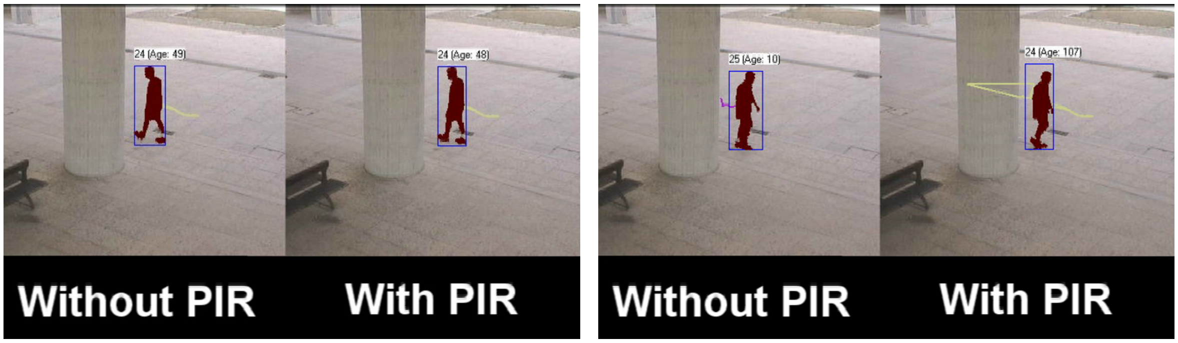
(g) Background - frame 1076 (h) Background - frame 2155 (i) Background - frame 2156

Fig. 12. Sensor-guided background update.

VII. Conclusions

Distributed surveillance is a challenging task for which a robust solution, working on a 7/24 basis, is still missing. This paper is meant to propose an innovative solution that integrates cameras and PIR (Passive InfraRed) sensors. The proposed multi-modal sensor network exploits simple outputs from the PIR sensor nodes (detecting the presence and the direction of movements of people in the scene) to improve the accuracy of the vision subsystem.

Two case studies are reported. In the first, the vision system, based on background suppression, fails due to a door that is opened. Since background is not immediately updated, the door is detected as a moving object (resolution is not sufficient to enable a correct motion detection). In this case, a PIR sensor is used to discriminate between the opened door and a real moving person. In the second case study, a person changes its direction when it is occluded by a column. The vision tracking algorithm relies on the constancy of the speed during occlusions and thus fails. A pair of PIR sensors are, instead,



(a) Before occlusion (b) After occlusion

Fig. 13. Consistent labelling after an occlusion exploiting a PIR node to detect direction changes

used to detect the change in direction and alerting the vision system.

The reported results demonstrate that using the integration between PIR sensors and cameras the accuracy can significantly be increased. Quantitative and extensive experimental results are underway to better evaluate the accuracy of the proposed system.

REFERENCES

[1] R. Valera Espina and S. Velastin, "Intelligent distributed surveillance systems: A review," *IEE Proceedings - Vision, Image and Signal Processing*, vol. 152, no. 2, pp. 192–204, Apr. 2005.
[2] http://www.smarthome.com/7527MC.HTML.
[3] D. Culler, D. Estrin, and M. Srivastava, "Guest editors' introduction: Overview of sensor networks," *IEEE Computer*, vol. 37, no. 8, pp. 41–49, Aug. 2004.
[4] http://www.xbow.com/Products /productsdetails.aspx?sid=3.
[5] G. Foresti, C. Micheloni, L. Snidaro, P. Remagnino, and T. Ellis, "Active video-based surveillance system," *IEEE Signal Processing Magazine*, pp. 25–37, Mar. 2005.
[6] F. Porikli and A. Divakaran, "Multi-camera calibration, object tracking and query generation," *Proc. of IEEE Intl Conference on Multimedia and Expo*, vol. 1, no. 1, pp. 653–656, July 2003.
[7] J. Krumm, S. Harris, B. Meyers, B. Brumitt, M. Hale, and S. Shafer, "Multi-camera multi-person tracking for easyliving," in *Proc. of IEEE Intl Workshop on Visual Surveillance*, 2000, pp. 3–10.
[8] J. Li, C. Chua, and Y. Ho, "Color based multiple people tracking," in *Proc. of IEEE Intl Conf. on Control, Automation, Robotics and Vision*, vol. 1, 2002, pp. 309–314.
[9] A. Mittal and L. Davis, "Unified multi-camera detection and tracking using region-matching," in *Proc. of IEEE Workshop on Multi-Object Tracking*, 2001, pp. 3–10.
[10] Z. Yue, S. Zhou, and R. Chellappa, "Robust two-camera tracking using homography," in *Proc. of IEEE Intl Conf. on Acoustics, Speech, and Signal Processing*, vol. 3, 2004, pp. 1–4.
[11] C. Jaynes, "Multi-view calibration from motion planar trajectory," *Image Vis. Comput.*, vol. 22, no. 7, pp. 535–550, July 2004.
[12] S. Khan and M. Shah, "Consistent labeling of tracked objects in multiple cameras with overlapping fields of view," *IEEE Trans. on PAMI*, vol. 25, no. 10, pp. 1355–1360, Oct. 2003.
[13] http://www.kalatel.com.
[14] H.-W. B. P. Bryant, "Some applications of a motion detecting camera in remote environments," *Technical Report*, Feb. 2003.
[15] S. de Vlaam, "Object tracking in a multi sensor network," *Master Thesis*, Aug. 2004.
[16] S. Rajgarhia, F. Stann, and J. Heidemann, "Privacy-sensitive monitoring with a mix of ir sensors and cameras," *Proceedings of the Second International Workshop on Sensor and Actor Network Protocols and Applications*, pp. 21–29, Aug. 2004.
[17] A. Sekmen, M. Wilkes, and K. Kawamura, "An application of passive human-robot interaction: Human-tracking based on attention distraction," *IEEE Transaction on Systems, Man, and Cybernetics*, vol. 32, no. 2, pp. 248–259, Mar. 2002.
[18] Y. Lu, L. Zeng, and G. M. Bone, "Multisensor system for safer human-robot interaction," *IEEE International Conference on Robotics and Automation*, Apr. 2005.
[19] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, "Detecting moving objects, ghosts and shadows in video streams," *IEEE Trans. on PAMI*, vol. 25, no. 10, pp. 1337–1342, Oct. 2003.
[20] R. Cucchiara, C. Grana, A. Prati, and R. Vezzani, "Probabilistic posture classification for human behaviour analysis," *IEEE Trans. on Systems, Man, and Cybernetics - Part A*, vol. 35, no. 1, pp. 42–54, Jan. 2005.
[21] A. Calderara, A. Prati, R. Vezzani, and R. Cucchiara, "Consistent labeling for multi-camera object tracking," in *Proc. of IEEE Int'l Conference on Image Analysis and Processing*, 2005.
[22] R. Cucchiara, A. Prati, L. Benini, and E. Farella, "T_PARK: Ambient intelligence for security in public parks," *Proceedings of IEE International Workshop on Intelligent Environments, Special session on Ambient Intelligence*, June 2005.