T_PARK: AMBIENT INTELLIGENCE FOR SECURITY IN PUBLIC PARKS[*]

Rita Cucchiara, Andrea Prati
University of Modena and Reggio Emilia - Italy

Luca Benini, Elisabetta Farella
University of Bologna - Italy

Abstract

In this paper, we present joint research activities in computer vision and sensor networks for a distributed surveillance of urban parks. Distributed visual surveillance of urban environments is one of the most interesting scenario in Ambient Intelligence; in addition, the automated monitoring of public parks, often crowded by children and adults, is still a very difficult task due to the number of objects of interests. In this context, integrating the power of low cost sensors with the information provided by cameras can lead to a more reliable solution to people tracking in wide areas. Specifically, the deficiencies of one approach can be (at least partially) covered by the advantages of the other. The goal is to perform people tracking in parks (to achieve *trackable parks - T_Parks*), both in zones covered by overlapped cameras and also, thanks to sensors, in zones not covered by any camera. In this paper, we propose a new technique for multi-camera people tracking based on a learning phase to automatically calibrate pairs of cameras and to build Areas of Field Views (AoFoVs) in order to establish consistent labelling of people. In addition, sensor networks distributed at the borders of the AoFoV give an estimation of the probability of people overlapping, triggering specific algorithms of face detection or head counting to identify the single person. The research of T_Parks is part of a two-year Italian project called *LAICA*, intended to provide advanced services for citizens and public officers based of Ambient intelligence technologies.

1 Introduction

Distributed visual surveillance of urban environments is one of the most interesting scenario in Ambient Intelligence; it consists in processing in real-time multimedia data coming from distributed external sources (video, audio, and other signals) to automatically extract information about the monitored environment, understand the presence of objects of interests, infer the behaviours of these objects w.r.t. the environment and, possibly, react with defined policies, depending on the detected situation. Typically, distributed video surveillance is used in either vehicular/traffic or people surveillance. People surveillance is more interesting from the research's point of view, 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.

This project describes the implementation of ambient intelligence methodologies and technologies for the monitoring of public urban parks that have the characteristics to be often crowded by children and adults, typically playing, running, walking or sitting on benches. Moreover, the changes of illuminations and the presence of occlusions due to trees or poles, make the outdoor scenarios cluttered and complex, requiring multiple cameras partially overlapped to solve occlusions and to cover a wide area.

Camera-based systems can be considered a rich source of local intelligence that can be exploited on the ambient intelligence scenario, although their cost (and, moreover, environmental impact) and the limited field of view make the complete coverage of wide areas infeasible. In addition, computer vision techniques, though dramatically improved in the last years, can not solve all the problems in people detection and tracking. The state-of-the-art approaches, even if working with 3D calibrated cameras, still lack in robustness in many practical cases.

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 the activation of video cameras [8]. 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 costs and low-power sensor nodes can now be developed with on-board processing capabilities, reconfigurability and wireless connectivity [5, 9]. We aim at integrating a wireless network of PIR-based sensors with a traditional vision system to provide drastically improved (in accuracy and robustness) tracking.

In particular, enabling technologies in PIR sensor net-

[*] This work was supported by the project L.A.I.C.A. (Laboratorio di Ambient Intelligence per una Città Amica), funded by the Regione Emilia-Romagna, Italy, under the development project “Piano telematico regionale”.

works and new algorithms for tracking multiple and overlapped people observed by multiple cameras have been designed and developed. The goal is to provide people tracking in parks (to achieve *trackable parks*, therefore the contraction *T_Parks*), both in zones covered by overlapped cameras and also, thanks to PIR sensors, in zones not covered by any camera. The proposed technique of auto-calibration of pairs of cameras with the automated learning of Areas of Field Views (AoFoV) enables the tracking and the establishment of consistent labels of people merged in groups in the view of one camera, but disjoint in the view of another. At this point, sensor networks, distributed on the borders of the AoFoV, give an estimation of the probability of overlapped people, triggering specific algorithms of face detection or head counting to identify the single person.

In this paper, we describe the proposed solution by focusing on two complementary aspects: on the one side, new solutions to allow a precise tracking of people in multiple cameras, solving the problem of consistent labelling in the whole area of overlapped fields of view is proposed. On the other side, the design of new field-configurable PIR-based wireless sensor nodes, with on-board processing capabilities, is presented. These sensors can be straightforwardly used as a trigger for the computer vision system to mark potential human presences entering in a specific zone. More specifically, they can be used to provide speed and direction information, as well as an indirect indication on the target size, which could be, with an adequate training, exploited to estimate the number of people present in the scene.

2 Related works

Video surveillance requires the identification of the subjects in the scene. Moreover, if these subjects move, their identities must be preserved in order to analyze their behaviours. This task, called "*consistent labelling*", is particularly challenging in distributed video surveillance, since the identity must be maintained also when the subject moves from the field of view of a camera to another.

Adjacent cameras can have overlapped field of views or not. In the case of non-overlapped cameras the only feature utilizable for maintaining the identity of a moving subject is its visual appearance. 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. To cope with this problem, in [11] a training procedure is used to model (by means of colour histograms) the change in the appearance from one camera to another. This information is combined with other data on the positions of the cameras in a probabilistic way using a MAP (Maximum A Posteriori) framework. Other approaches are based on assumptions on the specific case, such as [12] and [10] for traffic surveillance. Considering known the transition time, the non-deformable shape of a vehicle, and the forced direction in highways, the multi-camera tracking problem becomes much simpler.

Some works that use only subject's appearance have been proposed also in the case of overlapped cameras [14, 15]. However, in the case of partially overlapped cameras, the best choice is to exploit (also) geometrical information. Geometry-based approaches can be further subdivided into calibrated and uncalibrated approaches. In [17], 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 [19] 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. Once an occlusion is detected, homography is used to estimate the track position in the occluded view, by using the last valid positions of the track in it and the current position of the track in the other view (properly warped in the occluded one by means of the transformation matrix).

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. A relevant example of these is the work of Khan and Shah [13]. Their approach is based on the computation of the so-called *Edges of Field of View* (EOFOV, hereinafter), i.e. the lines delimiting the field of view of each camera and, thus, defining the overlapped regions. Through a learning procedure in which a single track moves from one view to another, an automatic procedure computes these edges that are then exploited to keep consistent labels on the objects when they pass from one camera to the adjacent.

Our approach is a suitable modification 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.

Previous approaches that combine sensors and computer vision are mainly focused on robot navigation and localization. Inertial sensors are used for calibrating cameras [6] or for recovering 3D structure from images [16], while proximity sensors are used, in combination with cameras, for object recognition and localization [7], but not for distributed video surveillance.

3 Vision-based surveillance

As stated in the introduction, the vision-based systems alone are not able to solve all the practical issues in distributed surveillance. On the other hand, sensors (audio, PIR, radar, etc.) can not provide sufficient information to be able to detect, track and analyze people in a complex scene. In many cases, the deficiencies of one approach are (at least partially) covered by the advantages of the other. Therefore, the integration of the two approaches to distributed surveillance can be extremely successful.

This section will describe a novel approach to multi-camera object tracking based on computer vision.

3.1 Single camera people tracking

People detection and tracking by single cameras are now very accurate and fast. Many approaches have been proposed in the literature. Their schemes are often similar: first, perform motion detection by separating points belonging to still parts from points belonging to moving parts (by means of background suppression, frame differencing, 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 three main modules: segmentation, tracking, and scene understanding. The segmentation module aims at extracting the *visual objects*. The first step uses the background suppression 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 [2], 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 $T^t = \{T_1^t, ..., T_m^t\}$ and assigns to each track T_i^t a status label: *moving*, *stopped*, *new*, or *undetected*. An object is classified as *stopped* when it is detected as still in the current frame and in at least a certain number of previous frames.

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 [4].

The knowledge about VOs and their status is exploited by a selective background model [2] in order to be both reactive to background changes and robust to noise. 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 [3], capable to discriminate between four postures (standing, sitting, crouching, and laying) and, consequently, to detect, by means of a state transition graph, interesting events, such as a person's fall. The above-mentioned probabilistic tracking is particularly suitable for posture classification since it is capable to preserve the appearance also in the case of occlusions.

3.2 Multi-camera people tracking

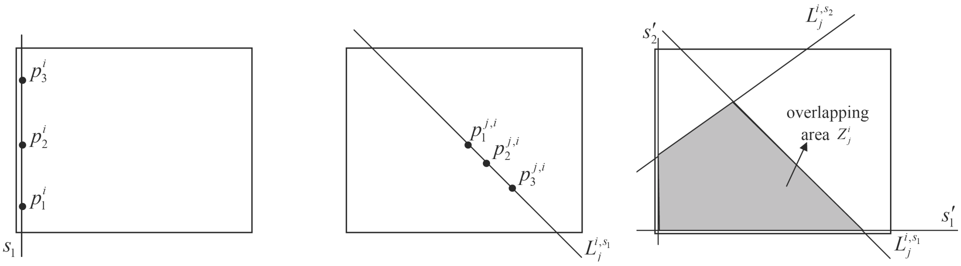
The consistent labelling problem has been already introduced in section 2. In that section, related works on calibrated or uncalibrated system has been also briefly described. Since in real installations often calibration can be tedious and imprecise (or even impossible), we propose to use a learning phase to automatically calibrate pairs of cameras.

The proposed solution starts from the creation of the so-called *edges of field of view* (EOFOV) to automatically calibrate cameras. Projecting the limits of the field of view (LOFOV) of a camera C^i on the ground plane ($Z = 0$), the so-called *3D FOV* can be obtained. In particular, they correspond to the intersection between the ground plane and the rectangular pyramid with its vertex at the camera optical center (the camera view frustum). Being $s \triangleq f_i(x, y) = 0$ one of the LOFOV of camera C^i defined by the equation f_i in its coordinate system, a 3D FOV line is denoted by $L^{i,s} \triangleq F_i(X, Y, Z) = 0$, where $Z = 0$ is one of the possible planes on which the LOFOV can be projected. In particular, the four 3D FOV lines $L^{i,s_h} \mid h = 1 \dots 4$ (where s_h corresponds to the image borders $x = 0$, $x = x_{max}$, $y = 0$, and $y = y_{max}$) can be computed. A projection of a 3D FOV line of camera C^i may be visible in another camera C^j partially overlapped with C^i. The FOV line (in 2D) of the line s of camera C^i seen by the camera C^j will be then denoted with $L_j^{i,s} \triangleq f_j(x', y') = 0$ and represents one of the EOFOV lines for the camera C^j (x' and y' are in the 2D coordinate system of C^j).

The EOFOV lines are created with a training procedure. A single person passes through at least two points of each limit of the FOV of two overlapped cameras. Given the constraint to have a single moving person, the support points (computed as the middle point of the bottom of the bounding box of the blob) in the two cameras can be matched and used to create the EOFOV line $L_j^{i,s}$ for the camera C^i. The equation of each line $L_j^{i,s}$ is computed by collecting a set of coordinates of the support point detected at the camera handoff and exploiting a Least Square optimization. Fig. 1 reports a sketch of process that, starting from the line s_1 of camera C^i, creates the EOFOV line $L_j^{i,s}$ in the camera C^j. Each line $L_j^{i,s}$ divides the image on camera C^j into two half-planes, one overlapped with camera C^i and the other disjoint. The intersection of the overlapped semi-planes defined by the EOFOV lines from camera C^i to camera C^j defines the Area of Field of View (AoFoV, hereinafter) Z_j^i (Fig. 1(c)).

To be sure that the matched points correspond to the person's support point, we delay the computation of the EOFOV lines to the moment in which the object is completely entered the scene of the new camera. This can bring to a displacement of the line with respect to the actual limit of the image, but it assures the correct match of the feet's position in the two views. As a consequence, the actual FOV lines are neither coincident nor parallel to the image border.

Thus, for two overlapped cameras C^i and C^j, the training procedure computes the AoFoV Z_j^i and Z_i^j. The four



(a) The LOFOV s in camera C^i (b) The EOFOV $L_j^{i,s}$ in camera C^j (c) The overlapping zone Z_j^i in camra C^j

Figure 1: Sketch of EOFOV creation.

corners of each of these areas defines a set of four points, $P_j^i = \{p_1^{i,j}, p_2^{i,j}, p_3^{i,j}, p_4^{i,j}\}$ and $P_i^j = \{p_1^{j,i}, p_2^{j,i}, p_3^{j,i}, p_4^{j,i}\}$, where the subscripts indicate corresponding points in the two cameras. These four associations between points of the camera C^i and points of the camera C^j on the same plane $z = 0$ are sufficient to compute the homography matrix H_j^i from camera C^i to camera C^j. Obviously, the matrix H_i^j can be easily obtained with the equation $H_i^j = (H_j^i)^{-1}$.

The approach proposed in [13] establishes the consistent labelling only in the exact moment of the camera handoff from C^i to C^j. This approach has two main limits: if two or more objects cross simultaneously (Fig. 2) an incorrect labelling can be established; or if they are merged from the view of C^j at the camera handoff, but then they separate, the consistent labelling with the labels of can not be recovered (Fig. 3).

With our approach, each time a new object is detected in the camera C^i in the overlapping area (not only at the moment of the camera handoff), 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. For the match we select the object in C^j whose support point is at the minimum distance in the 2D plane from these coordinates. The results achieved with this approach in the two cases above reported are shown in Fig. 2 and Fig. 3(c), respectively, where the correct label assignment is achieved.

In conclusion, our approach enables the correct establishment of the consistent labelling in the following cases:

1. only one person at a time crosses from camera C^i to C^j;
2. two or more people crosses from camera C^i to C^j, and they are disjoint and detected as separated by both cameras;
3. two or more people crosses from camera C^i to C^j, they are detected as merged by C^j, but as separated by C^i;
4. two or more people crosses from camera C^i to C^j, they are detected as merged by both cameras, but after some frames they separate (Fig. 3(c));

However, our approach is not able to handle the case 4 until objects get separated. For this aim, we should exploit alternative techniques for, at least, counting people during camera handoff. To this aim, computer vision alone can not be effective. For this reason, we plan to integrate proximity sensors into the system to provide an estimate of the number of people crossing from one camera to another.

4 Sensor nodes

In this section, we first describe the architecture and operation of the sensor nodes. Then, we provide experimental data demonstrating their capabilities beyond basic triggering.

4.1 Node architecture

A wireless sensor node is much more complex than a simple sensor. It packs together three sub-systems in a few cubic centimeters, namely: (i) the sensor, polarization and analog output conditioning circuits, (ii) the A2D conversion and digital processing unit, (iii) the radiofrequency (RF) transceiver and antenna. Moreover, the node contains power supply circuitry and an energy source (usually, a battery). In the following, we focus primarily on the first three subsystems.

4.1.1 The sensor

Our node is based on a passive infrared sensor. PIR sensors are solid-state devices which transduce incident infrared radiation into current. It is well known that objects irradiate infrared radiation depending on their temperature, hence PIR sensors can detect the perturbation in the infrared radiation spectrum caused by an object which is not at thermal equilibrium with the environment.

One of the most interesting features of these device is that the shape of the area of coverage can be accurately controlled by a suitably shaped Fresnel lens placed between the sensor and the environment. In most commercial PIR sensors, the area is a cone with elliptic base (squashed on the vertical dimension), and the height of the cone is a few meters (2-8 meters). The horizontal width (or equivalently, the aperture angle) can be modulated by choosing a suitable lens or by masking a part of the lens with infrared screening material (e.g. tin foil). Commercial PIR sensors are

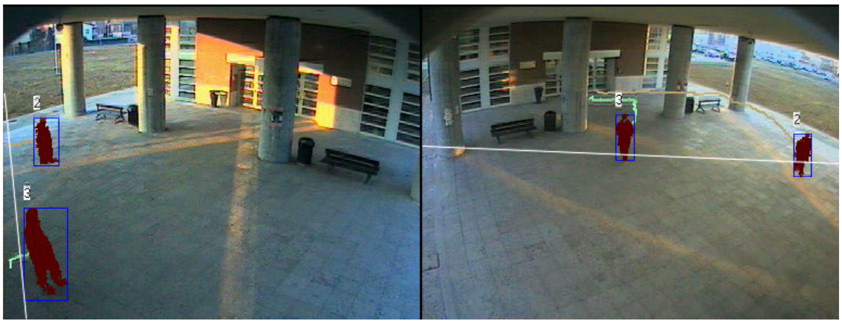
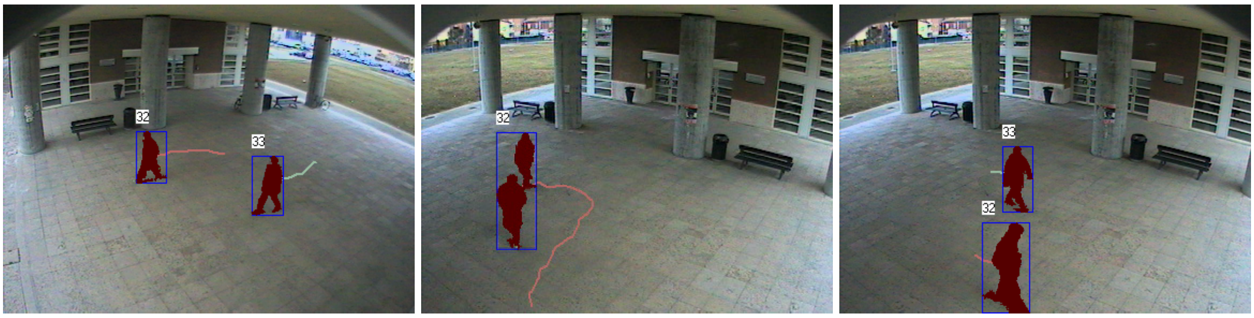


Figure 2: Example of simultaneous crossing of two objects.



(a) C^2 at frame #1250 (b) C^1 at frame #1250 (c) C^1 at frame #1260

Figure 3: Example of simultaneous crossing of two objects.

usually sold in pairs, with opposite polarization. Hence, an object moving along the alignment axis of the two sensors generates a pair of pulses (a negative and a positive pulse).

A very important characteristic of PIR sensors, which makes them highly suitable for wireless sensor network applications, is that they consume minimal energy. Their worst-case power consumption is less than 1mW. Note, however, that the output of the sensor must be amplified before digital to analog conversion. The amplification and sensor polarization circuits consume additional power which should be accounted for in power budgeting.

4.1.2 Digital processing

The node contains a low power digital processing block. Its main functions are to convert the output of the PIR to digital, and to process the incoming samples. Moreover, this block manages the wireless transmission protocol that grants wireless channel access to the various nodes in the network. Even though analog processing of the sensor output is most efficient in terms of components complexity and power, we aim at a flexible and configurable node architecture. For this reason, in our implementation we use a low-power ATMEGA8 RISC microcontroller by Atmel with integrated analog-to-digital converter. Its power consumption is 12mW in active mode, 4mW when idle and 0,1mW in power-down mode.

A key constraint on the selection of the digital components is power consumption, which should be around ten milliwatts during active data processing, and should drop to microwatt levels during idle periods. In order to aggressively reduce power dissipation, we developed a trigger-based power management solution which allows us to keep the digital components in low-power shutdown state unless the output of the sensor crosses an “event threshold”. When the threshold is crossed, an interrupt is raised for the processor, which starts to process the incoming sensor data.

4.1.3 Communication

Our nodes must be capable of short range (approximatively ten meters) wireless communication, as the data produced by the sensors and processed by the digital unit needs to be transmitted to the wired vision system for integration and coordination. Required data bandwidth is not very large: the sensor waveform sampled at Nyquist does not exceed 1200 bps; moreover, on-board data processing allows us to further reduce the required bandwidth (by feature extraction and compression).

Our wireless unit is based on a low-power RF transceiver, a digital transmission and receiver system implemented with TR1001 transceiver by RFM that operates in 868 MHz european free bandwidth. It uses OOK modulation with bit rate up to 100kbps. At this rate, multipath fading effects can be completely neglected up to distances of 50m.

Furthermore, we support a wireless media-access-control (MAC) protocol that allows time-shared channel access to several nodes in a star network architecture, where all sensor nodes communicate to a wired sensor network gateway, which collects the data and forwards it to the vision system through a standard USB connection. The MAC protocol and the gateway architecture are outside the scope of this paper. The interested reader is referred to [1, 18] for more details.

4.2 Experimental characterization

Our node prototype is shown in Figure 4. The node is implemented as a small-scale stacked multi-board system. Note that the architecture is highly modular, and one of the sub-systems can be changed by simply replacing the corresponding board. The size of the entire node is: $20 \times 20 \times 18mm$. Node power consumption is 66mW in the worst case when both the transceiver and the microcontroller are active, partitioned as follows: 6mW for the sensor block, 25mW for the microcontroller and 35mW for the transceiver. Average power is much lower since the sensor is active only when triggered by the threshold-crossing interrupt. Idle power is less than 10mW. Low cost off-the-shelf components have been used, and therefore the cost is minimal, even for prototypes. Total component cost for one node is around 20$.

Experiments where performed measuring the signal produced by the sensor when the same person crossed the coverage area at various distances from the sensor node. Figure 5(a) shows how the signal received by the PIR sensor decreases in amplitude as the distance increases. Note the saturation effect when the person is very close to the sensor. When the person is at a larger distance, the output peaks have smaller amplitude, but they are still very clearly discernible (the tests were performed in a room at 23^oC).

Interstingly, PIR sensors can provide much richer information. For instance, direction of movement. Figure 5(b) shows the signal detected from the 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). The signal detected in the first case is mainly a positive voltage, while the second signal is mainly a negative voltage. Thus, the time-domain output waveform can be easily processed to detect direction of movement.

Furthermore, we can use the sensor to obtain information on speed. Figure 6(a) shows the signal detected by the PIR sensor for a person moving at different speeds. As the speed increases the signal detected has a lower duration and amplitude, while its frequency increases. Simple frequency-domain analysis of the output waveform can therefore provide useful information on the speed of the movement.

Finally, the last plot in Figure 6(b) shows a significant difference between one person crossing the detection area and two or three people doing the same at the same time. Clearly, the output waveforms differ significantly in the three cases. This result demonstrates the possibility of providing useful information to the vision system on the number of people in a cluster.

Even though these preliminary results are promising, it is important to note that the output of the PIR sensor depends jointly on the temperature, the distance and the size of the object/person in the detection area. Thus, aliasing may be significant: similar waveforms could be generated by different combinations of the above parameters. As an example, consider two objects, one small, warm and close to the sensor and one bigger, colder and more distant from the PIR sensor (see leftmost side of Figure 7). In this case, the signal detected by the PIR sensor can be the same for the two objects. Using more than one sensor can help reducing aliasing, improving the robustness of the system as shown in the rightmost side of the same Figure. The bigger object is detected both by sensors a and b, thanks to its distance from both sensors, while the smaller object is detected only by sensor a. The same setup can be used to detect changes of direction inside the area detected by the array of sensors. In a similar manner, using a set of sensors placed along a vertical line, it is possible to distinguish among people and object of different heights. In general, matrix of sensors can help to better distinguish among different objects and to detect their shape. For these reasons multiple PIR sensors, connected in a wireless star-network to a data collection bridge, can provide a wealth of information to the vision system.

5 Implementation of T_Park

We report on a research part of a project called LAICA (Laboratorio di Ambient Intelligence per una Città Amica - Laboratory of Ambient Intelligence for a Friendly City) funded by the Regione Emilia-Romagna (Italy) and in collaboration with the municipality of Reggio Emilia, Italy, and several Italian universities and industrial companies. This multi-disciplinary project brings together the academic expertise and the industrial knowledge into several fields, from the low-power sensor networks, to the computer vision, to the middleware and mobile agents, to the communication. The objective of the project is the study and development of advanced services for the citizens and the public officers to improve personal safety and prevent crimes. These services will includes among the others: the automatic monitoring of pedestrian subways by means of mobile and low-power audio and proximity sensors; the automatic monitoring of traffic scenes by cameras for data collection and web-based delivery of traffic news to citizens; the generation of a feedback in pedestrian crossing systems to select the best duration of the green signal for the crossing; the automatic monitoring of urban parks with a plethora of cameras (both fixed and PTZ) and PIR sensors.

We focus on this last scenario. In order to provide extensive and repeatable experimentation of the system, we created a test bed on our campus, installing four partially overlapped cameras (three fixed and one PTZ) as sketched in Fig. 8, in a zone where many people are passing through, there are some benches and columns, thus trying to reproduce the conditions of an urban park. We are also installing PIR sensors, as indicated in Fig. 8.

The tests were carried out using a single camera probabilistic and appearance based tracking module [4]. EOFOV lines of the two cameras have been computed over a training video of 8000 frames. As an evidence of the goodness of the automatically obtained homography we report in Fig. 9 the mosaic image of two frames obtained merging a frame of a camera with a homographically distorted frame of the

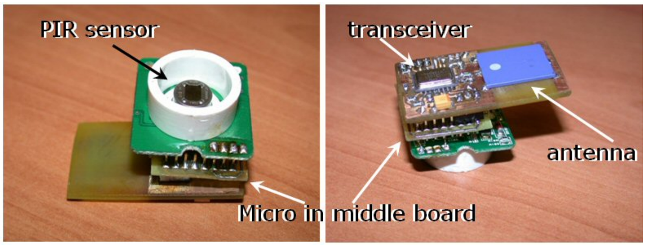
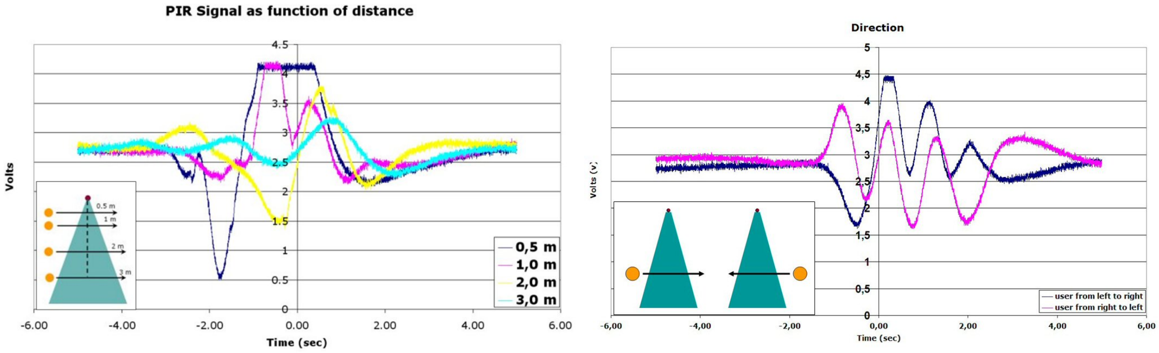


Figure 4: Prototype of the sensor node



(a) Function of the distance from the sensor node

(b) The same person walking from left to right and then from right to left

Figure 5: Voltage as function of the distance from the sensor and the direction of movement.

other camera.

	Sync. Trans.	Merged Trans.	N frames	N transitions	Correct	Incorrect
Video 1	No	No	8500	41	39	2
Video 2	No	No	3000	5	5	0
Video 3	Yes	No	1800	14	13	1
Video 4	Yes	Yes	2000	7	6	1
Video 5	Yes	Yes	500	2	2	0

Table 1: Experimental results

To test the consistent labeling algorithm, instead, we have tested the system not only in the simple conditions of the training phase, but also in presence of simultaneous transitions of more than one person at a time (Sync. Trans.) and in presence of transitions in which two people are merged (Merged Trans.) in a single track during the camera handoff and split far from the EOFOV.

In Table 1 we have reported the obtained results and some snapshots of the output of the system after the consistent labeling assignment are reported in Fig. 10. It is evident that the system has a very high accuracy. The incorrect matches are mainly due to errors in the lower modules, i.e. in the segmentation and the tracking algorithms, and to the case 4 reported in subsection 3.2. We are confident that, given the results reported in section 4, these situations can be handled with the help of PIR sensors.

6 Conclusions

This paper presents a proposal for the integration of sensor networks and cameras for distributed surveillance of urban parks. Even though we are still in a preliminary phase in which only few tests have been performed as proofs of concepts, the idea of integrating these two sources of data, trying to exploit the benefits of both, compensating the deficiencies, seems very promising.

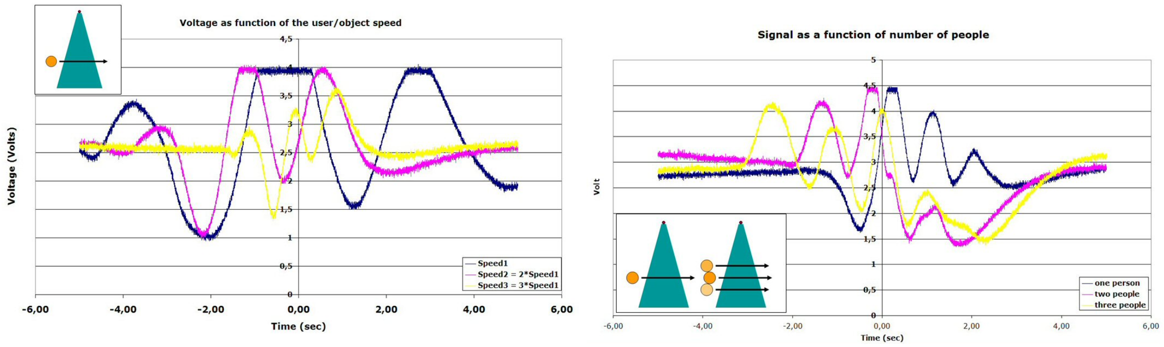
References

[1] M. Caccamo, L. Y. Zhang, L. Sha, and G. Buttazzo. An implicit prioritized access protocol for wireless sensor networks. In *Proc. of the IEEE Real-Time Systems Symposium*, pages 39–48, December 2002.

[2] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati. Detecting moving objects, ghosts and shadows in video streams. *IEEE Trans. on PAMI*, 25(10):1337–1342, October 2003.

[3] 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*, 35(1):42–54, January 2005.

[4] R. Cucchiara, C. Grana, G. Tardini, and R. Vezzani. Probabilistic people tracking for occlusion handling. In *Proc. of Int'l Conference on Pattern Recognition*, volume 1, pages 132–135, August 2004.



(a) The same person moving at different speeds

(b) Different number of people across the field

Figure 6: Voltage characterization in the presence of moving people.

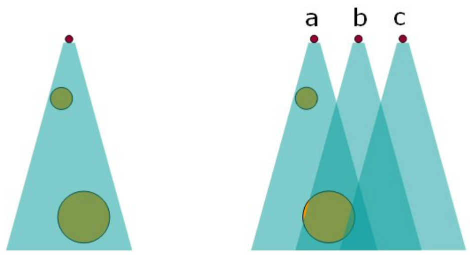


Figure 7: Usage of array of sensors

[5] D. Culler, D. Estrin, and M. Srivastava. Guest editors' introduction: Overview of sensor networks. *IEEE Computer*, 37(8):41–49, August 2004.

[6] J.-M. Frahm and R. Koch. Camera calibration with known rotation. In *Proc. of IEEE Intl Conference on Computer Vision*, volume 2, pages 1418–1425, October 2003.

[7] J. Hightower and G. Borriello. Location systems for ubiquitous computing. *IEEE Computer*, 34(8):57–66, August 2001.

[8] http://www.smarthome.com/7527MC.HTML.

[9] http://www.xbow.com/Products/productsdetails.aspx?sid=3.

[10] T. Huang and S. Russell. Object identification in a bayesian context. In *Proc. of Intl Joint Conf. on Artificial Intelligence*, pages 1276–1282, 1997.

[11] O. Javed, Z. Rasheed, K. Shafique, and M. Shah. Tracking across multiple cameras with disjoint views. In *Proc. of IEEE Intl Conference on Computer Vision*, volume 2, pages 952–957, 2003.

[12] V. Kettnaker and R. Zabih. Bayesian multi-camera surveillance. In *Proc. of IEEE Int'l Conference on Computer Vision and Pattern Recognition*, volume 2, pages 253–259, 1999.

[13] S. Khan and M. Shah. Consistent labeling of tracked objects in multiple cameras with overlapping fields of view. *IEEE Trans. on PAMI*, 25(10):1355–1360, October 2003.

[14] 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*, pages 3–10, 2000.

[15] J. Li, C.S. Chua, and Y.K. Ho. Color based multiple people tracking. In *Proc. of IEEE Intl Conf. on Control, Automation, Robotics and Vision*, volume 1, pages 309–314, 2002.

[16] J. Lobo and J. Dias. Vision and inertial sensor cooperation using gravity as a vertical reference. *IEEE Trans. on PAMI*, 25(12):1597 – 1608, December 2003.

[17] A. Mittal and L. Davis. Unified multi-camera detection and tracking using region-matching. In *Proc. of IEEE Workshop on Multi-Object Tracking*, pages 3–10, 2001.

[18] T. van Dam and K. Langendoen. An adaptive energy-efficient mac protocol for wireless sensor networks. In *Proc. of ACM Conf. on Embedded Networked Sensor Systems*, pages 171–180, November 2003.

[19] Z. Yue, S.K. Zhou, and R. Chellappa. Robust two-camera tracking using homography. In *Proc. of IEEE Intl Conf. on Acoustics, Speech, and Signal Processing*, volume 3, pages 1–4, 2004.

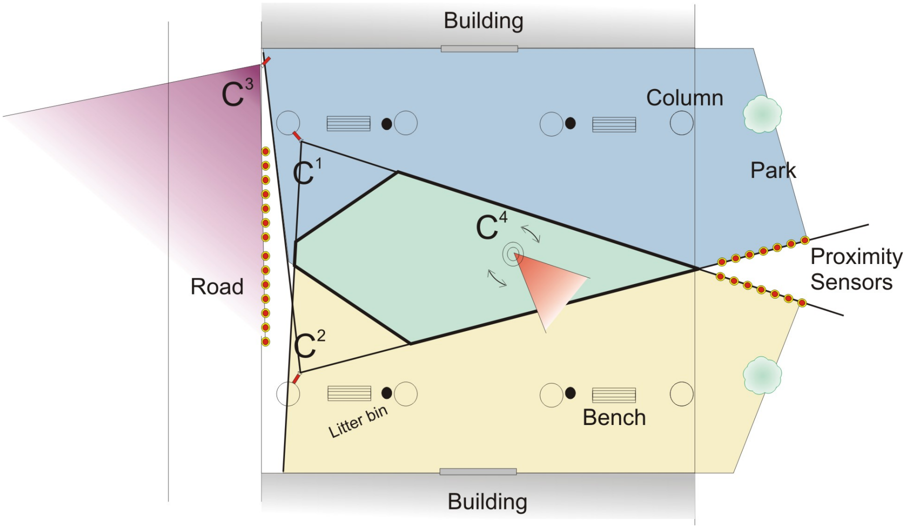


Figure 8: Map of our real setup.

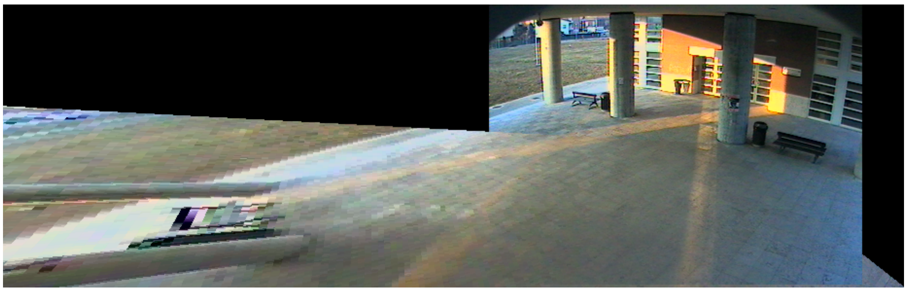
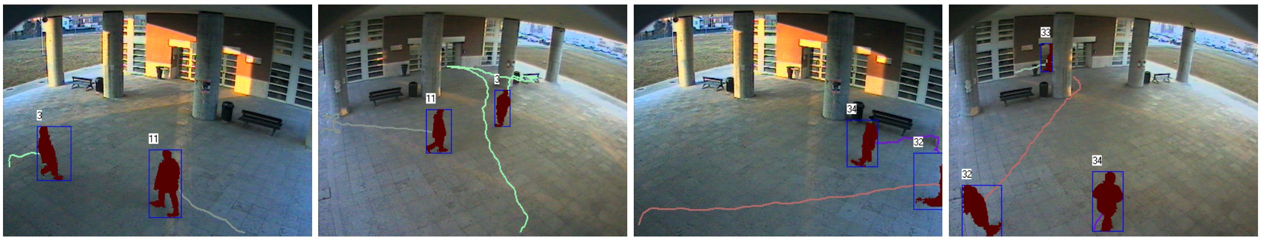


Figure 9: Automatically obtained mosaic image through homography.



(a) C^1 at frame #776

(b) C^2 at frame #776

(c) C^1 at frame #1490

(d) C^2 at frame #1490

Figure 10: Some snapshots of the output of the system after consistent labeling.