

Mutual Calibration of Camera Motes and RFIDs for People Localization and Identification

Rita Cucchiara
DII - University of Modena and
Reggio Emilia
Modena, Italy
rita.cucchiara@unimore.it

Michele Fornaciari
DII - University of Modena and
Reggio Emilia
Modena, Italy
45285@studenti.unimore.it

Andrea Prati
DiSMI - University of Modena
and Reggio Emilia
Reggio Emilia, Italy
andrea.prati@unimore.it

Paolo Santinelli
DII - University of Modena and
Reggio Emilia
Modena, Italy
paolo.santinelli@unimore.it

ABSTRACT

Achieving both localization and identification of people in a wide open area using only cameras can be a challenging task, which requires cross-cutting requirements: high resolution for identification, whereas low resolution for having a wide coverage of the localization. Consequently, this paper proposes the joint use of cameras (only devoted to localization) and RFID sensors (devoted to identification) with the final objective of detecting and localizing intruders. To ground the observations on a common coordinate system, a calibration procedure is defined. This procedure only demands a training phase with a single person moving in the scene holding a RFID tag. Although preliminary, the results demonstrate that this calibration is sufficiently accurate to be applied whenever different scenarios, where area of overlap between the field of view (FoV) of a camera and the “field of sense” (FoS) of a (blind) sensor must be efficiently determined.

Categories and Subject Descriptors

I.4.8 [Computing Methodologies]: Image Processing and Computer Vision—*Scene Analysis*; I.5.4 [Computing Methodologies]: Pattern Recognition—*Applications*

General Terms

Design, Algorithms, Security

Keywords

RFID technology, video surveillance, intruder detection, multi-sensor calibration

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ICDSC 2010 August 31 – September 4, 2010, Atlanta, GA, USA
Copyright 20XX ACM 978-1-4503-0317-0/10/08 ...\$10.00.

1. INTRODUCTION

There are several surveillance applications where *localizing* and *identifying* objects (people, vehicles, etc.) are crucial objectives for successive tasks such as collecting statistics, detecting intruders, analyzing paths of movement, recognizing suspicious behaviors, etc.. Localization and identification are typically two competing tasks since localization needs to not focus too much on a single object to have a wide coverage of the scene, whereas identification often requires a close-up of the object. This is particularly true in the case of cameras and when the objects are people: identification might require zooming on the person’s face and localization needs an unzoomed view to catch more people simultaneously and position them with respect to the scene. Moreover, cameras are typically mounted on high poles in order to reduce their number to cover the entire surveilled scene and the resolution is insufficient to recognize people by their face, or other biometric features (height, dresses, hair color). Finally, robust identification based on face recognition needs frontal or not occluded views and requires a certain degree of collaboration from the people themselves, which is unlikely to happen when looking for intruders. Alternative sensors can be used for localization and identification purposes. Among the many, RFID (Radio Frequency Identification) sensors [8] gained much attention thanks to their ease of use, low cost and touchless way-of-reading. The identification with RFIDs is accurate, using RFIDs for people identification introduces a limitation: RFID system detects only true positives, i.e. people wearing a RFID tag, but does not detect true negatives, i.e. “intruders” that are people not authorized or without the tag. Regarding localization, there have been previous attempts [19, 1, 10, 15] which use multiple RFID tags or readers to assess people’s locations using triangulation and RFID signal power (RSSI - Received Signal Strength Indicator). These approaches, however, do not guarantee a sufficient accuracy in the localization. With these premises, this work proposes to jointly use cameras and RFIDs to take the best from both of them: cameras are used to localize all the people in the scene (regardless if they are intruders or not), while RFIDs are used to identify allowed people only. A data fusion procedure could thus be employed to infer the number and location

of people present in the scene, their IDs and the presence of intruders. The scenario here is that of a wide open area with no designated entrances such as a construction working site, which makes the use of standard devices, such as badge, fingerprint or iris readers, unfeasible. In order to fuse the data coming from the camera and the RFID, camera calibration, at least partial, is indeed necessary to create a common coordinate system with the RFID sensors. To make the system desirable also in terms of costs, standard cameras are replaced with wireless sensors nodes (motes) equipped with low-cost CMOS cameras. These devices yield also the appreciable characteristics to be moveable and to not require wired power supplies, which are very useful for outdoor setups. Unfortunately, the use of a wireless network of sensors poses also several challenges, such as, for instance, communication and synchronization issues. From our perspective, however, the main problem is that, being the motes not fixed, standard calibration procedures are either not applicable or too time consuming. Thus, this paper addresses this last problem and presents a novel method to (semi-)automatically calibrate a joint RFID-camera system used for localization and identification of people in wide open areas. Despite the specific application scenario, this approach is extensible to different scenarios, where area of overlap between the field of view (FoV) of a camera and the “field of sense” (FoS) of a blind RFID sensor must be efficiently determined. The approach extends similar methodologies previously applied to multiple camera systems, using a person walking around equipped with a RFID tag as a “probe”. Data collected by heterogeneous sensors are processed together to find a precise and accurate area and allowing a mutual calibration. The paper is structured as follows: next Section will list previous works related to the use of RFIDs for localization and the joint use of cameras and RFIDs; Sections 3 and 4 will describe the general architecture and the devices used, while Section 5 will present the software modules used; Section 6 will detail the calibration procedure; Section 7 will report some results obtained and Section 8 will draw conclusions and future directions.

2. RELATED WORKS

Several methods for RFID tag localization have been proposed in the literature. They can be classified based on the operating environments, either indoor or outdoor. The Scout system proposed in [10] is an outdoor localization system based on off-the-shelf active RFID systems and a probabilistic localization algorithm. The system is capable to cover wide outdoor environments and it is organized in a hierarchical architecture comprising servers, RFID readers and RFID tags; object tags are used for object identification while so-called “reference tags” are used for the environmental parameter calibration, by measuring their RSSI information. The authors proved that in 90% of the localization estimation the system provides object location with an average error distance less than 7 meters. LANDMARC [15] is a prototypal system that uses RFID technology for locating objects inside buildings. In order to improve the accuracy, this system makes use of the reference tags as described above. The authors evaluated performance over a 35 square meters indoor lab using 4 readers and a grid of 16 reference tags. They have shown that the 50 percentile has an error distance of around 1 meter while the maximum error distance are less than 2 meters. The proposed approach

has some advantages: using extra cheaper tags there is no need for a large number of expensive RFID readers and the environmental dynamics can be easily accommodated. An Euclidean distance of the signal strength between a tracking tag and a reference tag is defined. For m reference tags each tracking tag has its E vector composed of the m distances. The proposed algorithm finds the unknown tracking tags’ nearest neighbours by comparing different E values. In [17] a RFID-based localization system for indoor localization of mobile robots is proposed. This system localizes RFID readers mounted on mobile robots using distributed tags on the floor. The reported experiments using a 5 x 4 tag square grid pattern with 7 cm inter-tag distance result in error variance reduction of the position estimation when power control is applied. The square root of the mean of error variance in distance is 5.6 cm in power control case and 5.9 cm in non-power control case. In all these methods based on RFIDs only the problem of “intruder” is not taken into account. The work in [16] is focused on a situation where a robot simultaneously interacts with two or more people and has to identify them with a passive-type RFID reader and floor sensors. A method is reported that integrates a person identification function provided by RFIDs and a person tracking function provided by floor sensors. To solve the association problem, if two or more people are around the robot, hypotheses are modeled using Bayesian networks and validated using the observations. Experimental results revealed that the developed system based on the proposed method successfully identified interacting people with 79.2% of the accuracy. In [6] a sensor fusion method for an heterogeneous sensor environment with visual and identification sensors is proposed. The proposed technique addresses the problem of the coverage uncertainty of the identification sensors. This problem is managed by grouping unassociated identifications. The comparison for association performance between the existing association methods and the proposed one in terms of number of identification sensors and tracking performance has shown the supremacy of the proposed approach in all the simulated cases. This work supposes to have a precise calibration between sensors available, without addressing the problem explicitly. Our approach could be employed in this situation.

3. THE CAMERA MOTES-RFID ARCHITECTURE

The platform we propose makes a joint use of cameras and RFID sensors for localizing and identifying people in wide open areas. The general hardware architecture is depicted in Fig. 1. The system is a heterogeneous sensor network where the main elements are the cameras (which in our case are supposed to be nodes of a wireless sensor network, i.e. camera motes), at least one RFID reader with its antenna, the active RFID tags and a server named “Base Station”. The camera motes are, generally speaking, able to establish wireless communication with each others and with the Base Station. The RFID reader is currently cable-connected to the Base Station server, but it can be equipped with a wireless interface becoming a node of the wireless network itself. The number of cameras that can be used is virtually unlimited, except for the inherent limit of the Base Station in terms of computational power and communication bandwidth. The use of multiple cameras is crucial to en-

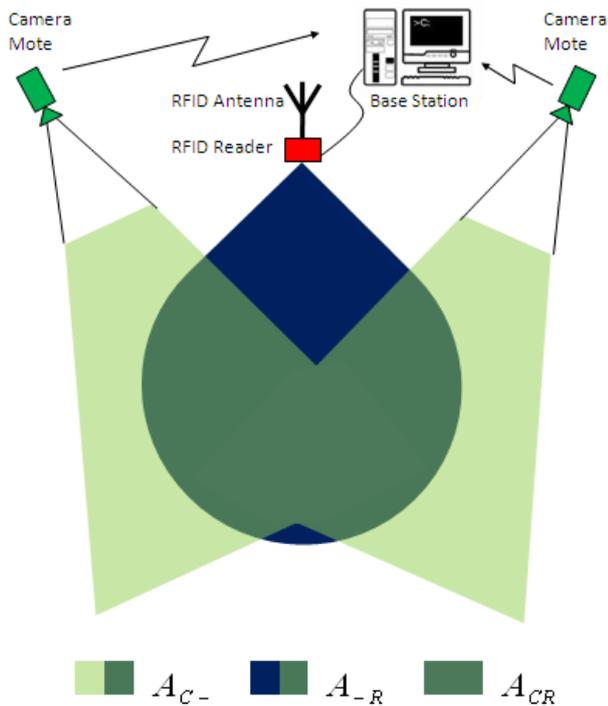


Figure 1: Sketch of the layout and the hardware architecture.

large the coverage of the scene. However, if the cameras are not coordinated the result of scene monitoring is much poorer. Conversely, if the label assigned by the video processing algorithms to a person can be kept consistent among the different camera views (procedure known as *consistent labeling*), the entire history of that person in the scene can be reconstructed: for instance, whenever the person is identified as an intruder, his/her movements in the whole scene can be back-traced. The FoVs of two close cameras can be either overlapped or disjoint. Often cameras' FoVs are disjoint, due to installation and cost constraints. In this case, the consistent labeling should be based on appearance only [13], i.e. matching two views of the same person only based on a more or less sophisticated model of the person's appearance. If the cameras' FoVs are (partially) overlapped, consistent labeling can exploit geometry-based computer vision. This could be done with a precise system calibration, but this is not often feasible, in particular if the cameras are pre-installed and intrinsic and extrinsic parameters are not available. Thus, partial calibration or self-calibration methods can be adopted to extract only some of the geometrical constraints, e.g. to compute the ground-plane homography [12, 4, 9], and base inter-view matching on distances on a common coordinate system. Similarly, the number of RFID readers depends on the coverage for identification that needs to be reached. The only requirement for merging identification with localization based on video processing is that the FoS of each reader is partially overlapped with the FoV of at least one camera. Therefore, in this general scenario, we can define the following areas (Fig. 1): A_{C-} , A_{-R} , A_{CR} , and two secondary areas $A_{\overline{CR}}$ and $A_{\overline{C\overline{R}}}$. The area A_{C-} corresponds to the cameras' FoV resulting from the merge of multiple cameras. A_{-R} is the RFID FoS, i.e. the area

where the RFID reader correctly detects tags. A_{CR} is the intersection of these two areas where both RFID-based identification and camera-based localization are possible. $A_{\overline{CR}}$ is the area where cameras cannot localize but the RFID can identify, that is $A_{\overline{CR}} = A_{-R} \setminus A_{C-}$, where \setminus represents the set difference. Similarly, $A_{\overline{C\overline{R}}}$ is the area where cameras can localize, but the RFID cannot identify, $A_{\overline{C\overline{R}}} = A_{C-} \setminus A_{-R}$. It is clear that the correct knowledge of these areas and their limits is crucial for inferring the status of the people in the scene. While cameras' FoVs can be determined using known methods such as the EoFoV (Edges of Field of View) [12] or the E²oFoV (Entry Edges of Field of View) [4], a method for calibrating the FoS of the RFID reader with respect to the camera is not available in the literature. Similarly to the E²oFoV where a single person is required to move in the scene in the calibration phase in order to collect pairs of points in two different overlapped views, our approach requires a single person to move in the scene of overlap between the FoS and the FoV: the pair image coordinates-power of the RFID signal is used to determine the area in which the identification of a person holding a RFID tag must be unmistakable.

4. THE HARDWARE ARCHITECTURE

The following subsections detail the camera motes and the RFID technology used in the described system, namely CITRIC wireless camera motes and active RFIDs.



Figure 2: CITRIC camera mote

CITRIC camera mote [5] consists of a camera sensor board running embedded Linux and a TelosB [18] network board running TinyOS. The camera board integrates a 1.3 megapixel SXGA CMOS image sensor, a frequency-scalable (208 - 520 MHz) Marvell PXA270 32-bit fixed point microprocessor, 16 MB FLASH, and 32 MB RAM memory. The general-purpose processor running embedded Linux allows the use of many well-studied image processing and computer vision algorithms coded in C/C++ such as the OpenCV library [2]. The TelosB network board uses the IEEE 802.15.4 protocol to communicate between camera nodes and the base station. The maximum data rate of 802.15.4 is 250 kbps per frequency channel (16 channels available in the 2.4 GHz band), far too low for a camera mote to stream images at a high enough quality and frame-rate for real-time applications. A key tenet of camera motes is to push computing out to the edge of the network and only send post-processed

data (for instance, low-dimensional features extracted from an image) in real-time back to the Base Station for further processing. A small number of functional blocks are in the camera board in order to minimize size, power consumption, and manufacturing costs. The measured power consumption of the camera daughter board running Linux with no active processes and with the mote attached, is from 527 to 594mW depending on the processor speed [5]. The measured power consumption running a typical background subtraction function is 970mW. This value is obtained at the processor speed of 520 MHz utilizing all the components of the camera, by both running the CPU and using the mote to transmit the image coordinates of the foreground objects. RFID is a technology for automated identification of objects and people, and may be viewed as a means of explicitly labeling objects to facilitate their “perception” by computing devices. A RFID system has several basic components including RFID readers, antennas and tags, and the communication between them. The RFID reader can read data emitted from RFID tags through an antenna. Readers and tags use a defined radio frequency and a protocol to transmit and receive data. Moreover, the tags can be divided in passive or active. Passive tags operate without a own power supply source: all the energy they need for the operation is provided by the reader and is collected by the passive tag antenna. Passive tags are much lighter and less expensive than active tags, offering a virtually unlimited operational lifetime. However, their read range is very limited. Active tags contain both a radio transceiver and a battery to power the transceiver. Since there is an on-board radio on the tag, active tags have a wider range than passive tags. With the capability of providing RSSI information, current advanced RFID systems have become a potential candidate for mass localization, though they are not enough accurate for our scopes. Our system uses off-the-shelf long range active RFID systems from IDENTEC SOLUTIONS [11]. This system works at 868 MHz frequency range, with a read range up to 100 meters, and includes active tags i-B, reader i-PortM, and an antenna. The active tags continuously send their unique ID at regular intervals (programmable ping rate) without being requested by a reader.

5. SOFTWARE ARCHITECTURE

Considering the hardware setup described in the previous section, several algorithms are needed to reach the goal. First, the cameras modules must process computer vision algorithms which detect and track moving objects in the scene. Specifically, in the case of camera motes, these algorithms should be computationally inexpensive due to the limited resources, but, at the same time, should guarantee a good robustness to noise, illumination changes, background camouflage, etc.. While these algorithms aim at computing people locations, the identities of allowed people are directly provided by the RFID technology. However, due to the unavoidable signal noise and to other sources of disturbance (e.g., presence of metal pieces, radio interferences, etc.), the RSSI obtained by the reader needs to be pre-processed to be cleaned. Finally, given both the set of people seen along the time by the camera motes and their corresponding locations, and the set of IDs sensed by the RFID reader, a data fusion strategy is required. It is exploited in the calibration step to automatically determine the area of overlap of the camera’s FoV and the RFID’s FoS; similar data will be used

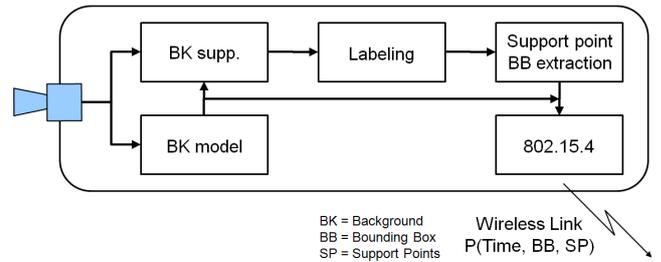


Figure 3: Data flow of the algorithms implemented in the camera mote.

after calibration to infer precisely the list and locations of intruders.

5.1 People Localization with Network of Cameras

The final prototype we aim at developing will make use of camera motes for the reasons reported in the introduction. However, we conducted also some tests using standard hard-wired cameras. In this case, cutting-edge algorithms for people detection and tracking can be used, such as the Sakbot approach [7], which is based on a sophisticated background suppression method and an appearance-based tracking. Unfortunately, in the case of camera motes, both the computational and the communication resources are limited and thus several operations must be performed on-board and only high-level post-processed data are sent to the Base Station. In order to reduce the amount of base station data processing each camera mote (Fig. 3) performs background subtraction and blob identification (by grouping all connected regions in the foreground image), and compute a set of boxes bounding the resulting blobs. Then, for each image the full set of geometrical information extracted from the bounding boxes, such as its dimensions and position in the image plane, are sent to the Base Station to be used for tracking moving objects. Because of the reduced computing capabilities of the camera motes the computational complexity of the algorithms is a crucial factor which justifies the choice of the simple background suppression technique. In this approach, each incoming frame is compared with the background model, which is updated by using median filtering as described in [5]. A reduction in the number of false alarms in outdoor environments is achieved by carefully tuning the blending factor value used to update the background model and discarding small bounding boxes (with area below the a given threshold). To enable data consistency and coordination, each camera motes send local time information. To keep consistency information, camera motes time synchronization is required [14]. On the Base Station a set of remote commands are available in order to aid the camera motes deployment. During the people localization activities, camera motes send to the base station geometrical information related to the moving objects detected on each frame. Simultaneous tracking of multiple objects using Kalman filter [3] and consistent labeling among different views (using the HECOL approach [4]) are performed by the Base Station.

5.2 People Identification with RFID Technology

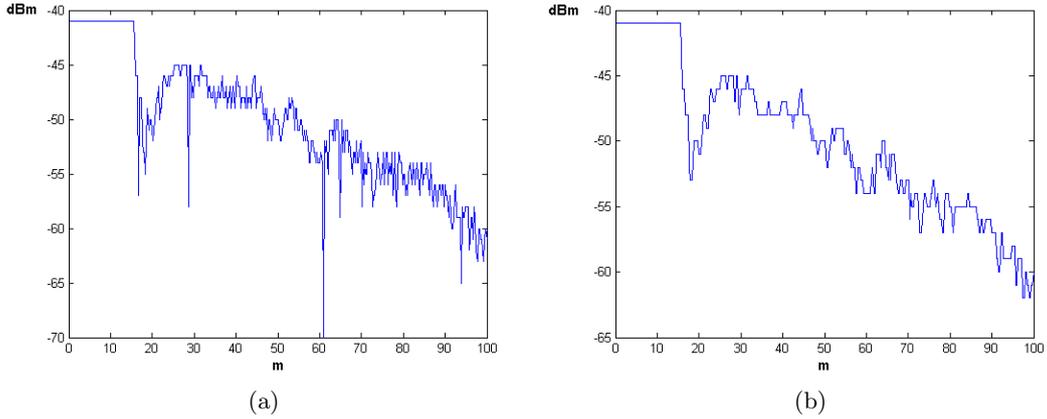


Figure 4: RFID signal strength RSSI over distance: (a) before pre-processing, (b) after applying median filtering.

It is well known that the signal provided by RFID technology is very noisy, in particular when using active tags which have a very high reading range. Fig. 4(a) reports an example of the RSSI value (expressed in dBm, i.e. power ratio in decibels (dB) referenced to one milliwatt) for a person moving slowly away from the RFID antenna. It is evident from this plot that the RSSI generally decreases, but with a lot of noise and with a particular but not accidental drop around 20 meters from the antenna. Given these problems of not linearity between RSSI and distance, both the localization and the identification using directly the RSSI information are doomed to fail.

Thus, we first clean the signal applying a median filter of width W (in our experiments $W = 5$ - result in Fig. 4(b)) and we must select a threshold to decide whether the RFID is sensed/identified or not. Hereinafter, we will consider the tag in the status “sensed” when the RSSI is above or equal to the threshold, and “not sensed” otherwise¹. In the case of a single person moving in the scene, the status of each reading can be automatically associated with the location of the single blob detected. Fig. 5 shows some experiments obtained changing the threshold. With higher threshold, as for instance Fig. 5(a) with respect to Fig. 5(d), the number of locations with a sensed RFID (in green) is lower and is higher the number of detected location where the RFID is not sensed (in red).

6. CALIBRATION OF CAMERAS AND RFIDS

Similar to what has been done for calibrating pairs of overlapped cameras [4], our proposal for calibrating the RFID reader with a camera is to consider a procedure where a single person holding a RFID tag is moving in the scene. This simplification, aside from easing the detection and tracking by means of the camera, allows a straightforward association between the (x, y) coordinates of the lower support point of the person (extracted as described in Section 5.1) and the corresponding RSSI value of the tag. Fig. 6 shows an example of the tracking (Fig. 6(a)) and the corresponding plot of

the collected points with the color indicating the RSSI value (Fig. 6(b)).

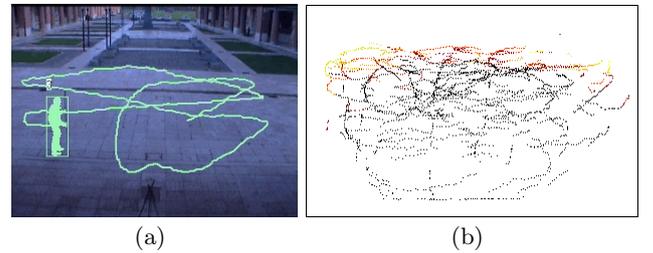


Figure 6: Example of the calibration procedure: (a) example of tracking and localization with the camera, (b) plot of locations with color indicating the RSSI value.

By starting from the distribution of points of Fig. 6(b) and applying the thresholding described in the previous section, two sets of points/locations can be obtained: above and below the threshold. Considering that the ping rate of the RFID reader is not so high (1 sample/second in our case) and that a person should not correspond to a single point but has a certain occupancy in the scene, we apply on these two sets a morphological dilation with a circular structuring element (with radius equal to 10), resulting in two sets called DAT (Dilated Above Threshold) and DBT (Dilated Below Threshold). These two sets for the example in Fig. 6(b) are reported in Fig. 7(a) and 7(b), respectively.

It is worth noting that after dilation these two sets are not disjoint and their intersection defines an area $U = DAT \cap DBT$ where the RFID reading can not be considered certain (i.e., it may happen or not that a tag is sensed). Similarly, the two areas $CS = DAT \setminus U$ and $CN = \Omega \setminus DAT$ represent the “certainly sensed” and “certainly not sensed” areas, respectively, where Ω represents the domain of the locations (i.e., the image). By applying morphological operations to these two areas and using a convex hull to close polygons, we can obtain three different areas (Fig. 7(c)): the so-called area $NEAR$ is defined by the pixels inside the convex hull applied on CS (inside yellow area in Fig. 7(c)); the area FAR by the pixels outside the convex hull on DAT (out-

¹Physically the tag is sensed also under the threshold but with a low RSSI.

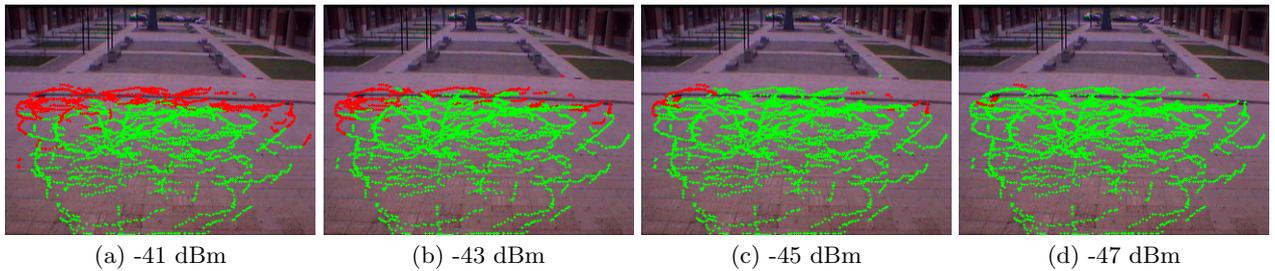


Figure 5: Distribution of below-threshold (red) and above-threshold points by changing the threshold (in dBm).

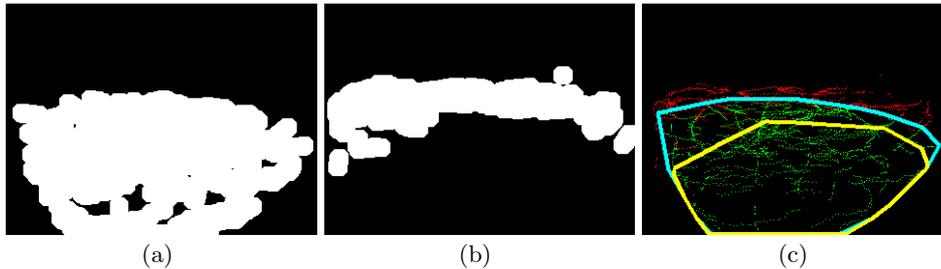


Figure 7: Example of the calibration procedure: (a) *DAT* area for example of Fig. 6(b), (b) *DBT*, (c) final determination of the areas.

side cyan area in Fig. 7(c)); finally, the area *UNCERTAIN* which is contained between the *NEAR* and the *FAR* areas. An evaluation of the accuracy in determining these areas will be reported in Section 7.

The calibration of these areas can be then exploited to assert, for any person detected and tracked by the cameras, whether he/she is in the *NEAR*, *UNCERTAIN* or *FAR* area, as an input of more or less sophisticated reasoning engines. As a possible way of proceeding, consider Fig. 8 where all the possible transitions C_i of a person (Fig. 8(a)) and all possible transitions R_i of a RFID tag (Fig. 8(b)) are shown. Thanks to the calibration, it is possible to use the area *NEAR* as FoS of the RFID, and, considering the movement of an allowed person (thus holding a tag), some of the combinations between C_i and R_i are not compatible. Table 1 summarizes these combinations. Note that some combinations with $R3$ are yet compatible but of no use for the inference (indicated with C^*). These considerations can be a starting point for at least discarding not compatible observations of cameras and RFIDs. It is, however, not enough to infer the presence of intruders and more sophisticated approaches, based on statistical inference, such as Transferable Belief Model (TBM) [20] should be used.

7. EXPERIMENTAL RESULTS

The work reported in this paper is part of a larger project for the monitoring and security of construction working sites, which includes the detection of intruders. At this early stage, since the inference engine has not been yet fully developed, the preliminary results only aim at evaluating the accuracy of the calibration process. To achieve this, we consider two outdoor scenarios, one taken with a single normal camera, and one taken from a pair of camera motes (Fig. 9). In the case of the normal high-resolution camera, it has

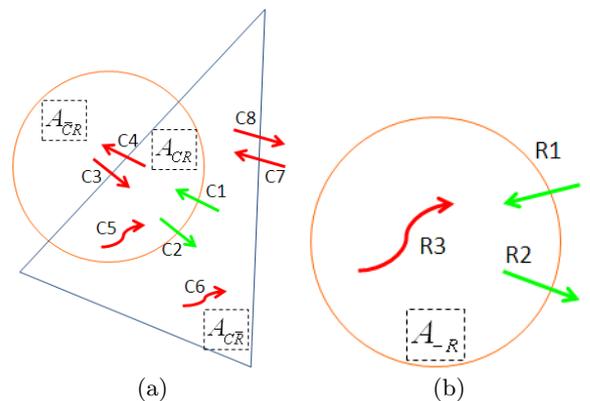


Figure 8: Example of transitions in the FoV of the camera (a) and in the FoS of the RFID (b).

been connected to the Base Station through a wired connection, the video feeds are transmitted to the Base Station as is, and the Sakbot algorithms (Section 5.1) are run on the Base Station directly. In the case of the pair of camera motes, instead, the communication is wireless, the algorithms are much lighter (see again Section 5.1) and directly implemented on board. Then, only the locations of the moving people are sent to the Base Station which implements tracking, consistent labeling among the two camera motes and data fusion with the RFID. The layout used in our experiment with two camera motes is shown in Fig. 10.

The experimental methodology is as follows. A video with a single person holding a RFID tag is used as training for delimiting the *NEAR* and *FAR* areas as described above. Then, a certain number of videos with multiple people holding tags is used as a testing set: in our experiments 4 videos

		R1	R2	R3
		$A_{\overline{R}} \rightarrow A_{-R}$	$A_{-R} \rightarrow A_{\overline{R}}$	$A_{-R} \rightarrow A_{-R}$
C1	$A_{\overline{C\overline{R}}} \rightarrow A_{C\overline{R}}$	C	NC	NC
C2	$A_{C\overline{R}} \rightarrow A_{\overline{C\overline{R}}}$	NC	C	NC
C3	$A_{\overline{C\overline{R}}} \rightarrow A_{C\overline{R}}$	NC	NC	C*
C4	$A_{C\overline{R}} \rightarrow A_{\overline{C\overline{R}}}$	NC	NC	C*
C5	$A_{C\overline{R}} \rightarrow A_{C\overline{R}}$	NC	NC	C*
C6	$A_{\overline{C\overline{R}}} \rightarrow A_{\overline{C\overline{R}}}$	NC	NC	NC
C7	$A_{\overline{C\overline{R}}} \rightarrow A_{C\overline{R}}$	NC	NC	NC
C8	$A_{C\overline{R}} \rightarrow A_{\overline{C\overline{R}}}$	NC	NC	NC

Table 1: Possible combinations of transitions of camera tracks (row) and RFID tags (column). C=Compatible, NC=Not Compatible. Please note that $A_{\overline{R}}$ represents the area outside the RFID’s FoS, while $A_{\overline{C\overline{R}}}$ is composed of points outside both the FoV and the FoS. C* means that the combination is yet compatible but of no use for the inference.

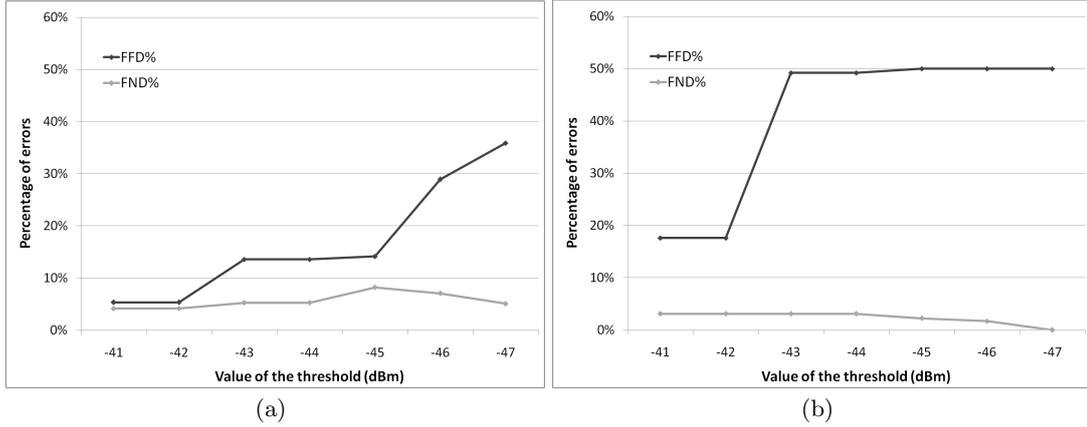


Figure 11: False Near Detections (FND) and False Far Detections (FFD) in percentage for the scenario with a normal camera (a) and with a pair of camera motes (b).

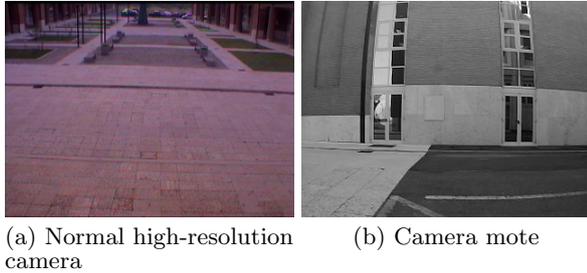


Figure 9: Snapshots of the two scenarios.

for the first scenario and 12 videos for the second have been used. With respect to the testing videos, the number of *false near detections* FND (i.e., the number of locations which are supposed to be in the *NEAR* area but in which the RSSI value is below the threshold - the same used during the training phase) and the number of *false far detections* FFD (i.e., the number of locations which are supposed to be in the *FAR* area but in which the RSSI is still above the threshold) are employed to evaluate the accuracy of the calibration procedure. These two values, averaged on the testing videos and with different values of the threshold, are reported for the two scenarios in Fig. 11. As expected, the FND parameter is low because of the high signal to noise

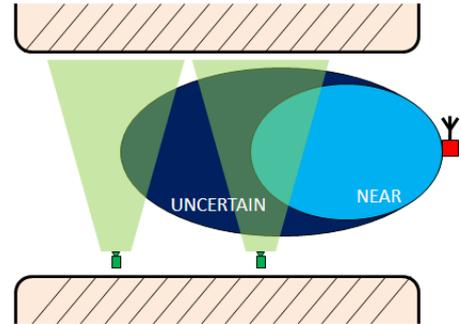


Figure 10: Layout of the scenario with two camera motes.

ratio of the points in the *NEAR* area. Viceversa reducing the threshold value increases the number of points having a lower signal to noise ratio justifying the increase of the FFD value.

8. CONCLUSIONS

This paper presents an easy calibration procedure for (semi-)automatically determining the area of overlap between the cameras’ FoV and the RFID’s FoS. The system exploits cam-

era notes to detect, track and localize moving people and active RFID technology for identifying allowed people. In order to move towards an inference engine capable to detect and localize intruders in wide open areas (with no designated entrances), a calibration of the two subsystem is presented.

Future directions include a more comprehensive experimentation in order to better validate the calibration procedure and the implementation of an inference engine based on the Transferable Belief Model.

9. REFERENCES

- [1] M. Bouet and A. dos Santos. Rfid tags: Positioning principles and localization techniques. In *Wireless Days, 2008. WD '08. 1st IFIP*, pages 1–5, nov. 2008.
- [2] D. G. R. Bradski and A. Kaehler. *Learningopencv, 1st edition*. O'Reilly Media, Inc., 2008.
- [3] R. Bucy and P. Joseph. *Filtering for Stochastic Processes with Applications to Guidance*. AMS Chelsea Publ., 2005.
- [4] S. Calderara, A. Prati, and R. Cucchiara. Hecol: Homography and epipolar-based consistent labeling for outdoor park surveillance. *Computer Vision and Image Understanding*, 111(1):21–42, 2008. Special Issue on Intelligent Visual Surveillance (IEEE).
- [5] P. Chen, P. Ahammad, C. Boyer, S.-I. Huang, L. Lin, E. Lobaton, M. Meingast, S. Oh, S. Wang, P. Yan, A. Yang, C. Yeo, L.-C. Chang, J. Tygar, and S. Sastry. Citric: A low-bandwidth wireless camera network platform. *Distributed Smart Cameras, 2008. ICDCS 2008. Second ACM/IEEE International Conference on*, pages 1–10, sept. 2008.
- [6] S. H. Cho, S. Hong, and Y. Nam. Association and identification in heterogeneous sensors environment with coverage uncertainty. In *AVSS '09: Proceedings of the 2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance*, pages 553–558, Washington, DC, USA, 2009. IEEE Computer Society.
- [7] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati. Detecting moving objects, ghosts and shadows in video streams. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(10):1337–1342, Oct. 2003.
- [8] K. Finkenzer. *RFID Handbook: Fundamentals and Applications in Contactless Smart Cards and Identification*. John Wiley & Sons, Inc., New York, NY, USA, 2003.
- [9] W. Hu, X. Xiao, Z. Fu, D. Xie, T. Tan, and S. Maybank. A system for learning statistical motion patterns. 28(9):1450–1464, September 2006.
- [10] X. Huang, R. Janaswamy, and A. Ganz. Scout: Outdoor localization using active rfid technology. *Broadband Communications, Networks and Systems, 2006. BROADNETS 2006. 3rd International Conference on*, pages 1–10, oct. 2006.
- [11] IDENTEC SOLUTIONS AG. www.identecsolutions.com.
- [12] S. Khan and M. Shah. Consistent labeling of tracked objects in multiple cameras with overlapping fields of view. 25(10):1355–1360, Oct. 2003.
- [13] C. Madden, E. D. Cheng, and M. Piccardi. Tracking people across disjoint camera views by an illumination-tolerant appearance representation. *Mach. Vis. Appl.*, 18(3-4):233–247, 2007.
- [14] M. Maroti, M. Maroti, B. Kusy, B. Kusy, G. Simon, G. Simon, A. Ledeczi, and A. Ledeczi. The flooding time synchronization protocol. pages 39–49. ACM Press, 2004.
- [15] L. M. Ni, Y. Liu, Y. C. Lau, and A. P. Patil. Landmarc: indoor location sensing using active rfid. *Wirel. Netw.*, 10(6):701–710, 2004.
- [16] K. Nohara, T. Tajika, M. Shiomi, T. Kanda, H. Ishiguro, and N. Hagita. Integrating passive rfid tag and person tracking for social interaction in daily life. *Robot and Human Interactive Communication, 2008. RO-MAN 2008. The 17th IEEE International Symposium on*, pages 545–552, aug. 2008.
- [17] Y. Park, J. W. Lee, and S. Kim. Improving position estimation on rfid tag floor localization using rfid reader transmission power control. In *ROBIO '09: Proceedings of the 2008 IEEE International Conference on Robotics and Biomimetics*, pages 1716–1721, Washington, DC, USA, 2009. IEEE Computer Society.
- [18] J. Polastre, R. Szewczyk, and D. Culler. Telos: enabling ultra-low power wireless research. In *IPSN '05: Proceedings of the 4th international symposium on Information processing in sensor networks*, page 48, Piscataway, NJ, USA, 2005. IEEE Press.
- [19] T. Sanpechuda and L. Kovavisaruch. A review of rfid localization: Applications and techniques. In *Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, 2008. ECTI-CON 2008. 5th International Conference on*, volume 2, pages 769–772, may 2008.
- [20] P. Smets. Data fusion in the transferable belief model. In *Information Fusion, 2000. FUSION 2000. Proceedings of the Third International Conference on*, volume 1, pages PS21–PS33 vol.1, july 2000.