A Reasoning Engine for Intruders' Localization in Wide Open Areas using a Network of Cameras and RFIDs

Rita Cucchiara[1], Michele Fornaciari[1], Razia Haider[1], Federica Mandreoli[1],
Riccardo Martoglia[1], Andrea Prati[2], Simona Sassatelli[1]
[1]DII, [2]DiSMI - University of Modena and Reggio Emilia - Italy

Abstract

Wide open areas represent challenging scenarios for surveillance systems, since sensory data can be affected by noise, uncertainty, and distractors. Therefore, the tasks of localizing and identifying targets (e.g., people) in such environments suggest to go beyond the use of camera-only deployments. In this paper, we propose an innovative system relying on the joint use of cameras and RFIDs, allowing us to "map" RFID tags to people detected by cameras and, thus, highlighting potential intruders. To this end, sophisticated filtering techniques preserve the uncertainty of data and overcome the heterogeneity of sensors, while an evidential fusion architecture, based on Transferable Belief Model, combines the two sources of information and manages conflict between them. The conducted experimental evaluation shows very promising results.

1. Introduction

Wide open areas surveillance, especially for people monitoring, is a challenging research field. The first difficulty relies on the need of multiple sensors in order to cover "wide" areas. This calls for data fusion methodologies to handle noise and data redundancy. Moreover, "open" areas are normally unconstrained without obliged entrances or defined paths. These conditions make the surveillance job much more difficult, especially when the goal is to provide both localization and identification of targets, and specifically people. A typical example of application is the real-time control of people in order to detect and localize *intruders* [20]. Actually, the two tasks of identifying and localizing intruders in wide open areas are competing when only video analysis is adopted: identification might require zooming on the person's face and localization needs an unzoomed view to find the correct position with respect to the scene, even if many people are potentially present.

For this reason, in this scenario the data fusion of information coming from a network of cameras should be enriched with other information acquired by different sensors. Among the many alternative sensors, the RFID technology gained much attention thanks to its ease of use, low cost and touch-less way-of-reading. RFID technology enables applications to identify people carrying small RFID tags in an environment equipped with RFID readers. Therefore, the joint use of cameras and RFIDs could make exploitable the best from both of them: camera-based systems can localize all the people in the scene (regardless if they are intruders or not), while RFIDs can identify allowed people only. In this scenario, an intruder is any person which is localized by cameras but not identified by RFID readers (thus, potentially not holding any tag). The two tasks of localization and identification are certainly made more challenging when sensors (both cameras and RFIDs) are affected by noise, uncertainty, distractors and complex scenarios: illumination changes, occlusions and reflexes can make the task for computer vision algorithms applied to cameras hard, while multiple signal sources and the presence of metallic objects can introduce much noise in RFID signals.

This is the first proposal of joint use of cameras and RFIDS in real noisy and complex wide open areas for intruder localization. We propose a new architecture and specific algorithms system for intruder detection relying on:

- sophisticated filtering techniques for singular sensor modality that preserve the uncertainty of data in the form of probabilities and overcome the heterogeneity of sensors through the introduction of common locations which the data coming both from cameras and RFIDs are mapped to;
- an evidential fusion architecture, based on Transferable Belief Model - TBM [17], that processes uncertain data, combines the two sources of information and manages conflicts between them in order to "map" RFID tags to people detected from the cameras, thus highlighting potential intruders.

2. Related Works

In the last years, the surveillance of wide open areas has become an urgent matter for security reasons. In particular, the so-called "third generation" of intelligent video

surveillance systems [20] has been conceived to provide more accurate information by fusing more sensors, possibly belonging to different types (not only cameras). As stated in [20], this requirement poses several challenges, summarized in: (i) distributed versus centralised intelligence, (ii) data fusion, (iii) probabilistic reasoning framework, (iv) multi-camera surveillance techniques.

Our work proposes an integrated framework with joint use of RFID sensors and cameras for detecting intruders. The following related works will basically focus on: (a) examples of joint use of RFIDs and cameras for surveillance applications; (b) management of RFID sensors; (c) data fusion using a probabilistic reasoning framework.

Though distributed video surveillance is not new by itself, the use of different sensors and advanced reasoning techniques is not so diffused in the literature. For instance, the combination of cameras and RFID sensors is proposed in [22, 23] only to avoid to expose the privacy of authorized people in recording video streams. In this case, however, the scenario comprises buildings with doors and entrances where the user is forced to authenticate by means of RFID technology before entering in the monitored area: if the person is authorized the recorded video is protected by a watermarking algorithm. Therefore, this approach does not allow to identify those who are authorized and those who are not among several people.

Another example is provided by the work in [13] where a robot simultaneously interacts with two or more people and has to identify them with a passive-type RFID reader and 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. In [4] a sensor fusion method for an heterogeneous sensor environment with visual and identification sensors is proposed. The problem of the coverage uncertainty of the sensors is managed by grouping unassociated identifications. Despite these examples and to the best of our knowledge, a complete RFID/camera system for intruder detection in wide open area is still missing in the literature.

Coming to the use of RFID sensors, in last few decades, RFID technology has emerged significantly with many real time applications, such as product tracking and assess management, object and people authentication, health care etc. Nevertheless, the data management in these RFID applications poses a number of challenges [3]. In particular, the nature of an RFID data stream is noisy, redundant and unreliable, making it unsuitable for direct use in applications: among the issues that need to be effectively faced in most RFID deployments, the most common are conflicting readings (a tag is read by multiple antennas in conflicting ways) and missed readings (readers commonly detect only about 60%-70% of the tags in their range) [16]. For all these reasons, the unreliable data streams must be transformed into precise, reliable streams that can be meaningful to applications. Several techniques have been proposed for the analysis and processing of raw noisy RFID data [15]. A number of techniques propose to clean the data streams deterministically. For instance, [8] proposes a declarative framework for RFID data cleaning and processing which makes use of a window-based adaptive smoothing filter, producing more reliable RFID data streams by interpolating missed readings. Other techniques, instead, exploit the probabilistic nature of RFID data and manage their inherent uncertainty in the form of probabilities and correlations, so to achieve even higher effectiveness in the application scenarios they are applied to [16, 19, 12]. For instance, [16, 11] generate probabilistic streams by inference on a Hidden Markov Model (HMM). Then, probabilistic inference is required in order to extract high-level complex events from the low-level atomic events acquired by the readings. For example, in tracking applications, the location of the objects is unknown to the system and observed low level sensor data is translated into precise and more reliable estimates about the location of these objects [16, 12]. Note that all such RFID systems define locations on the basis of actual places/areas which are of interest to the final users (e.g. a restricted-access room), as reflected also by the supported queries and the produced results (e.g. "Find out which rooms entered Paul today"). In this regard, the combined RFID/camera system we propose differs from this vision: as we will see in the next sections, in our case the subdivision of the open area in a number of locations is solely an internal parameter which can be fine-tuned by the system administrator so to allow the best possible (and effective) communication between the RFID sensors' and the cameras' processing engines. While our final users do not even need to be aware of the chosen locations, they will surely benefit from a "smart" location choice, allowing the system to better and faster identify/localize the people.

Finally, our approach makes use of a Transferable Belief Model (TBM) for inferring the mapping between people and RFID tags. TBM has been used in the literature for different applications, such as for the classification of the camera motion [9] and for developing a system for advanced driver assistance [5]. The work in [5] is particularly interesting since it considers two heterogeneous sources of information (omnidirectional cameras and a laser scanner), but with similar objective (the localization of vehicles). Here, instead, the two heterogeneous sources also have heterogeneous purposes.

3. System Description

The high-level scheme of the proposed system is shown in Fig. 1. The two sources of information are the current frame f_t provided by the distributed cameras and analyzed by the *Video processing* module, and the RFID signals r_t

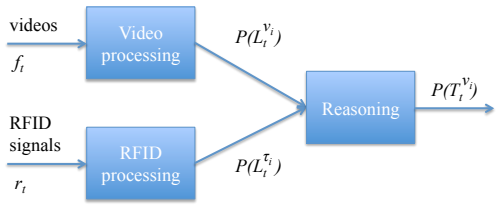


Figure 1. High level system description

provided by the tags and elaborated by the *RFID processing* module.

In the remainder of the paper we will refer to random variables using the uppercase letter and to single value with the lowercase letter. Specifically, we will refer to the following entities:

- $\mathcal{V}$ as the visual objects (typically people) detected by the video processing module, $\mathcal{V} = (\nu_{i=1,\dots,n})$;
- $\mathcal{T}$ as the tags deployed to the authorized personnel, which can be then identified by the RFID processing module, $\mathcal{T} = (\tau_{i=1,\dots,s})$;
- $\mathcal{L}$ as the locations in the scene, $\mathcal{L} = (\lambda_{i=1,\dots,k})$. Locations are used to correlate data coming from cameras with those coming from RFID.

The video processing module computes the set of visual objects at time t and estimates the probability distribution $P\left(L_t^{\nu_i}\right)$ of the random variable $L_t^{\nu_i}$ over the set of locations $\mathcal{L}$, one for each visual object ν_i. In other words, for each location λ_j, $P\left(L_t^{\nu_i} = \lambda_j\right)$ represents the probability that object ν_i is in λ_j.

Since a wide area can be monitored using multiple cameras only, our system makes use of a sophisticated algorithm for *consistent labeling* in partially-overlapped fields of view. The videos acquired by each camera are processed using the Sakbot (Statistical And Knowledge Based Object deTector) system [7]. This motion detection algorithm is specifically designed to ensure a robust and reliable background estimation even in complex outdoor scenarios and is based on constructing a background model by means of a pixel-by-pixel temporal median model with a selective knowledge-based update stage. Once people are detected, they are tracked along time using a probabilistic appearance-based tracking algorithm [21] that takes into account not only the status vector containing position and speed, but also the memory appearance model and the probabilistic mask of the person's shape. Finally, consistent labeling among different cameras is obtained using the homography-based approach presented in [2].

Similarly, the RFID processing module will elaborate the raw RFID signal r_t to estimate the probability distribution $P\left(L_t^{\tau_i}\right)$ of the random variable $L_t^{\tau_i}$ over the set of locations $\mathcal{L}$, one for each tag τ_i.

Finally, the reasoning module takes the two probability distributions, $P\left(L_t^{\nu_i}\right)$ and $P\left(L_t^{\tau_i}\right)$, as input and, for each visual object ν_i, it outputs the probabilities of carrying any of the tags in $\mathcal{T}$ together with the probability of carrying none of them. Formally, it outputs the random variable $T_t^{\nu_i}$ whose probability distribution is defined over the set of tags $\Omega = \mathcal{T} \cup \{\mathbb{k}\}$, where $\mathbb{k}$ is the dummy tag virtually held by an intruder. The higher is $P\left(L_t^{\nu_i} = \mathbb{k}\right)$ the higher is the probability that the visual object ν_i is an intruder.

In the following, we will first focus on the RFID processing module and then on the reasoning module. Finally, we will make some considerations on the role of locations in the system.

3.1. RFID Processing

The *RFID Processing* module infers the values of $L_t^{\tau_i}$ by exploiting a Hidden Markov Model (HMM) [14]. Indeed, HMMs are usually applied in contexts where the attributes of interest are not directly observable to infer their state on the basis of other related observed data. The module uses an HMM to produce, at each timestamp, a distribution over each tag location (i.e. the *hidden variables* or *states*) based on *observations* that, being sensor readings, include four types of information: 1) the identifier of the tag the reading is concerning to; 2) the identifier of the antenna the tag is seen by; 3) the timestamp of the reading; 4) the Received Signal Strength Indicator (RSSI) of the reading. Nevertheless, the main important feature of this kind of models is that they allow to combine prior domain knowledge about the system behavior with the actual observations to compute the most likely values of the hidden variables. While observations are directly evaluable, the prior knowledge about the system is represented by conditional probability distributions (CPD) which are referenced as the parameters of the HMM. Specifically, the parameters of our HMM are the following:

- the *initial states probability* $P\left(L_0\right)$: it encodes the knowledge about the initial state of the system and is usually assumed to be a uniform distribution among all the possible locations in Λ;
- the *transition probability* $P\left(L_t|L_{t-1}\right)$: it encodes the knowledge about how the state at time t depends on the state at time $t-1$[1]. It is modeled as a matrix whose rows ad columns are associated to the available locations so that each cell $[i, j]$ contains the probability value of having a movement from location i to location j (as an example, if two locations are separated by a wall the corresponding cell contains the value 0).

[1]Note that, according to the well-known Markov principle, these models typically assume that the variables at time t directly depend on the variables at time t and $t-1$ only and, hence, two consecutive time instances are sufficient for completely representing the whole system.

- the *observation probability* $P(R_t|L_t)$: it encodes the knowledge about how the observation at time t depends on the state at time t; this information is typically not available and, thus, has to be learned from training data. To this end, we adopt the popular statistical method called *Maximum Likelihood Estimation (MLE)* which, given learning data, estimates the value of the probability function parameter that maximizes the likelihood of the observed data (i.e. that makes the learning data "most likely"). Actually, MLE allows us to compute the conjunctive probability $P(R_t, L_t)$, from which observation probability $P(R_t|L_t)$ can be easily computed by applying the Bayes theorem.

Nevertheless, the final aim of modelling a stochastic process with an HMM is to obtain the *posterior probability distribution* $P(L_t^{\tau_i})$ over the hidden variable $L_t^{\tau_i}$ given the observed measurements. This task is called "inference" and different algorithms can be used to this purpose. Among the others, we decide to exploit a popular Monte Carlo algorithm called *Particle Filtering* [1], usually adopted in sample-based inference processes. The algorithm works by computing and constantly maintaining sets of particles to describe the historical and present states of the model. Fig. 2 represents a schema of the steps executed by the algorithm at each time instant t. Specifically, given the observed values $r_t^{\tau_i}$ for each identified tag τ_i, the algorithm works by iteratively executing the following steps:

Initialization: during this phase, an initial set of particles is created by randomly sampling from the initial states probability $P(L_0)$.

Prediction: during this phase, the state of hidden variables at time t is estimated by using their state at time $t-1$ and exploiting the parameters of the HMM. More precisely, for each existing particle p_{t-1}^i at time $t-1$ a new particle p_t^i is created for time t by sampling from $P(L_t|L_{t-1})$.

Filtering: in this phase, the observations r_t arrived at time t are used to update the states previously estimated for time t. More precisely, each particle p_t^i is assigned a weight based on the values of the observed variables at time t and on the observation probability $P(R_t|L_t)$. This weight is proportional to $P(R_t = r_t^{\tau_i}|L_t = \lambda)$ where λ is the location of p_t^i.

Re-sampling: in this phase, the particles created in the *Filtering* step are re-sampled in order to generate a new set of particles, all with the same weight. This task is necessary in order to avoid degeneracy, i.e. the case where a single particle has all the weight.

Broadly speaking, each particle p_t^i represents a guess about the location of tag τ_i. Then, after a number of iterations, the inference task is performed: to compute the posterior probability $P(L_t^{\tau_i})$ we can indeed simply count the number of particles in each location and divide it by the total number.

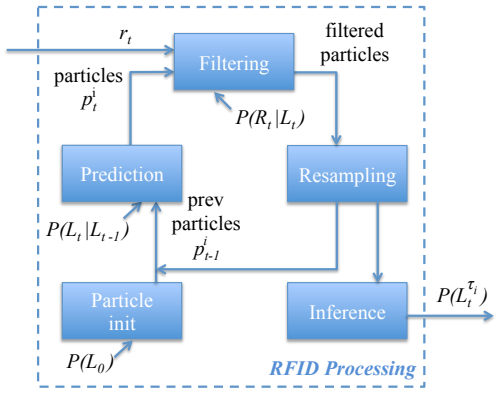


Figure 2. Detail of RFID processing schema

3.2. Reasoning Engine

The reasoning engine has the main objective (see Fig. 1) to fuse the inference coming from video and RFID processing modules by means of the Transferable Belief Model (TBM) [17].

TBM is a model that represents quantified belief (or weighted opinions) held by a "belief holder", called *System* hereinafter, based on the belief function theory. Given a general *frame of discernment* $\Delta = \{H_1, \ldots, H_b\}$ containing b mutually and exhaustive hypothesis related to a given problem (*closed world assumption*), belief can be represented by a *basic belief assignment* (*bba*), which is a function $m : 2^{\Delta} \rightarrow [0, 1]$ that satisfies $\sum_{A:A\subseteq\Delta} m(A) = 1$ and assigns a value in $[0, 1]$ to each subset $A \subseteq \Delta$ representing the part of System's belief that is allocated to the hypothesis A. The advantage of the TBM over the classical Bayesian approach resides in its ability to represent every state of partial beliefs: total ignorance ($m(\Delta) = 1$), partial ignorance and total knowledge ($m(A) = 1$). It is a powerful model to deal with *uncertainty*, which may results from sensor noise, misreading or semantic noise.

In order to map tags with people, the reasoning engine goes through the following five main steps.

Step 1 – Belief on locations. The probabilities over the locations of tags, $P(L_t^{\tau_i})$, and visual objects (or people, in our case), $P(L_t^{\nu_i})$, are translated in a Bayesian belief function [18] on the frame of discernment $\Delta \equiv \mathcal{L}$, which represents the set of locations of the scene. Therefore, for the tag τ_i (the same for the person ν_i) at time t, the resulting *bba* is: $m_t(\lambda_j) = P(L_t^{\tau_i} = \lambda_j)\ , \forall \lambda_j \in \mathcal{L}$, where the subscript t has been added to indicate time and for congruency with previous notation. Each of these new data provided by sensors is then used to update System's knowledge about localization until time $t-1$, encoded as a set of belief functions (one for each tag and each person). The new information about τ_i (resp. ν_i) updates only the belief function for that particular tag (resp. person). The updating task is performed using an appropriate combination rule. Among

others [18], we use Dubois-Prade's conjunctive combination rule because it merges coherent information in a conjunctive way and conflicting ones in a disjunctive way:

$$m_{1\sqcap 2}(A) = \sum_{X\cap Y=A} m_1(X)m_2(Y) + \sum_{\substack{W\cap Z=\emptyset \\ W\cup Z=A}} m_1(W)m_2(Z)$$
$$m_{1\sqcap 2}(\emptyset) = 0 \quad (1)$$

where $m_1 = m_{t-1}$, $m_2 = m_t$ and $A, X, Y, W, Z \subseteq \mathcal{L}$.

Step 2 – Similarity between locations. It is worth noting that the more the localization of tags and people is accurate, the higher is the mass for the same (set of) location(s) of a tag and its holder. Therefore, the comparison between the beliefs on localization of tag τ_j (m_{τ_j}) with the one of person ν_i (m_{ν_i}) returns a similarity value that indicates the support to the decision of mapping τ_j to ν_i derived from all information available to System at this moment. Defining $fe(x)$ as the set of focal elements (i.e. a subset $A \subseteq \mathcal{L}$ where $m_x(A) > 0$) of the belief function relative to a generic x, we use a measure which accounts for the similarity between focal elements through the Jaccard index [10]:

$$\psi(\nu_i, \tau_j) = \sum_{A\in fe(\nu_i)} \sum_{B\in fe(\tau_j)} m_t^{\nu_i}(A)\cdot m_t^{\tau_j}(B)\cdot \frac{|A\cap B|}{|A\cup B|} \quad (2)$$

Step 3 – Evidence generation. The above mentioned similarity values are exploited to generate a new piece of information (*evidence*) encoded as a *bba* on the frame of discernment $\Delta \equiv \Omega = \mathcal{T} \cup \{\mathbb{k}\}$ relative to the person ν_i. Let $(\psi(\nu_i, \tau_1), \ldots, \psi(\nu_i, \tau_q))$ be the set of similarity values between the person ν_i and the q tags sensed at this moment (with $q \leq s$, where s is the total number of available tags, as defined in Section 3), ordered by decreasing value. We create focal elements of the evidence using the following criterion:

$$m\left(\Gamma_j\right) = \begin{cases} \psi\left(\nu_i, \tau_{j-1}\right) - \psi\left(\nu_i, \tau_j\right) & \text{, if } j < q \\ \psi\left(\nu_i, \tau_j\right) & \text{, if } j = q \end{cases} \quad (3)$$

where $\Gamma_j = \bigcup_{1\leq i\leq j,\ j\leq q} \tau_i$.

Moreover, to equal the sum of masses to 1 and considering that a person could be an intruder, we define respectively:

$$m(\Omega) = (1 - \psi(\nu_i, \tau_1)) * \beta \quad (4)$$
$$m(\mathbb{k}) = (1 - \psi(\nu_i, \tau_1)) * (1 - \beta) \quad (5)$$

with $\beta < 1$. A rule of thumb suggests $\beta = 0.7$.

Step 4 – Belief update and decision. For each person we combine the relative evidence with System's belief on the mapping between people and tag using again Eq. (1), but this time m_1 represents the System's belief function on the mapping for person ν_i, m_2 is the new evidence relative to ν_i, and $A, X, Y, W, Z \subseteq \Omega$. Because System has to choose which tag is held by each person ν_i, it constructs a probability function on Ω in order to make the optimal decision, using the following "pignistic transformation":

$$P\left(L_t^{\nu_i} = \omega\right) = \sum_{A:\omega\in A\subseteq\Omega} \frac{m(A)}{|A|(1 - m(\emptyset))} \quad (6)$$

where $\omega \in \Omega$ can be either a tag or $\mathbb{k}$. $P\left(L_t^{\nu_i} = \omega\right)$ denotes the probability that the tag ω is held by the person ν_i.

Step 5 – Decision and belief reinforcement. The latter information can be useful to strengthen the mapping. For each tag τ_j, we consider only the *maximum probability value* over all people: $mpv(\tau_j) = \max_i P(L_t^{\nu_i} = \tau_j)$. In other words, we assume that a tag τ_j is held only by the person which is more likely to hold it at time t. We thus generate, for the person having the $mpv(\tau_j)$, the *bba*: $m(\tau_j) = mpv(\tau_j)$ and $m(\Omega) = 1 - mpv(\tau_j)$. If a person does not hold any tag with the highest probability, he/she is considered as an intruder, because we assume that a tag can be held by only one person. For each of these persons ν_i we thus generate the *bba*: $m(\mathbb{k}) = P(L_t^{\nu_i} = \mathbb{k})$ and $m(\Omega) = 1 - P(L_t^{\nu_i} = \mathbb{k})$. Finally, using Eq. (1) we combine these evidences with the System's belief on the mapping for that person (therefore reinforcing its belief on that particular tag) and the negated evidence (Eq. (7) below) on the others. Again, we need to re-define the meaning of m_1 and m_2 of Eq. (1): in this case, m_1 is the System's belief function on the mapping for the considered person and m_2 is the evidence just generated for him/her. We define also a belief function negate operator consistent with the closed world assumption:

$$m(A) = \sum_{B\subseteq\Omega:\overline{B}=A} m(B) \quad (7)$$

where $\overline{B} = \begin{cases} \Omega - B & \text{, if } (|B| = 1) \text{ and } (B \neq \{\mathbb{k}\}) \\ \Omega & \text{, otherwise} \end{cases}$.
This operation is coherent with the fact that a tag can be held by one person only at a given time, except for the dummy tag $\mathbb{k}$.

TBM is then very suited to our application because (i) it supports the combination of different pieces of evidence coming from different sources, as the cameras and RFIDs in our case, (ii) it allows the information to be updated in time, (iii) it allows to propagate the uncertainty associated to the available information and (iv) knowledge is refined using only available information with no further assumption. The reasoning engine is thus capable to map tag and people, and to identify intruders when mapped with the dummy tag $\mathbb{k}$.

3.3. On the choice of locations

Now that we have described in detail the inner architecture of our combined RFID/camera system, we can shortly

discuss the criteria for choosing locations. The output of the system, consisting in the probability of a given tag being held by a given person, fully reflects our ultimate goal, i.e. associating the tags to the people so to be able to identify and locate possible intruders: locations are not part of the final output, even if a precise localization of each visual object is still of course possible thanks to the video analysis. Differently from most proposed RFID-only deployments, in our system the locations are not necessarily known to the final users and they do not necessarily have to coincide with actual places of interest. Instead, locations are the mean which data coming from both the camera and RFID processing modules are mapped to, thus location choice becomes an internal parameter which can be fine-tuned so to maximize the cooperation between the two modules and the final effectiveness of the system. This can be done by subdividing the open area so that: a) the resulting locations can still be correctly distinguishable by the RFID and video processing modules in most situations (e.g. sufficient size, disposition compatible with the deployed cameras/antennas configuration, and so on); b) the number of locations is sufficiently large to allow a substantial amount of location changes to be identified by the RFID and video modules, so that enough observations can be fed to the reasoning engine, which will in this way eventually provide more accurate results. The latter requirement could be easily satisfied by a preliminary offline analysis of the paths typically covered by people in the area, for instance by reviewing previously captured videos. As an alternative, an automatic method to determine the locations which maximize their identification by RFIDs can be employed (see for instance the work in [6]).

4. Experimental Results

For evaluating the effectiveness of the envisaged application we have conducted experiments in different challenging situations, consisting of real wide open areas in our Campus, where several cameras are installed. In these scenarios it would be almost impossible to achieve good results without a reasoning engine capable of dealing with imprecise, uncertain and missing data.

Fig. 3 shows the overview of the testbed: Fig. 3(a) shows a 3D reconstruction of the area with the cameras (indicated by a red arrow) and the antenna used in our tests properly highlighted; Fig. 3(b) reports a bird-eye view of the setup with the locations represented by bounded areas and the antenna indicated by a green arrow. Upon this challenging setup, we have collected data from cameras and RFID tags in two cases. In the first case (*Case 1*) only one camera and one RFID antenna (covering the whole area to be monitored thanks to the active technology of our RFIDs) have been used, dividing the area in four locations (from 1 to 4 in Fig. 3(b)): different scenes of increasing complexity (with more people and tags, and with an intruder) have been analyzed. In *Case 2*, four cameras and seven locations have been used, allowing a larger scene coverage (thanks to the consistent labeling techniques described in Section 3) and considering a more realistic scenario.

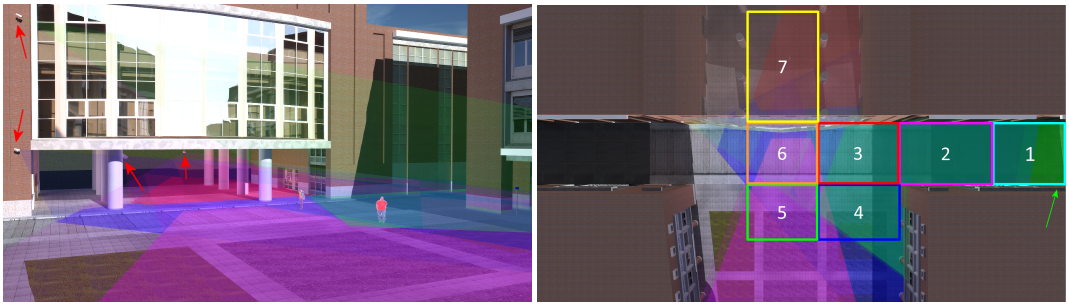
During the training phase, we have used a single person as a probe to collect RSSI samples from the tag in the different chosen locations, and then perform MLE on them in order to map the locations $\lambda_i \in \mathcal{L}$. During the testing phase, instead, particle filtering is applied to infer/track the location of the RFID tags. Particle filtering has been initialized with 500 particles where initial probability distribution for each location is uniform. Regarding the prediction, a uniform transition matrix has been defined according to a map of locations, e.g. the probability of moving from one location to others is uniform for all but the case of two locations which are not directly connected with each other or separated by some barrier (e.g. wall), where the probability is set to zero.

In the following we will report the descriptions and the results of the different tests performed, where $\Omega = \{A, B, C, D\} \cup \{\mathbb{k}\}$.

***Case 1*: two people, one authorized (tag A), the other intruder - Fig. 4**: in this case, the authorized people (ν_1) and the intruder (ν_2) walk side by side, making identification more challenging. Fig. 4(a) and Fig. 4(b) show the probability of different tags to be mapped to the respective person. Average precision and recall values (see caption of Fig. 4) are 100.0%.

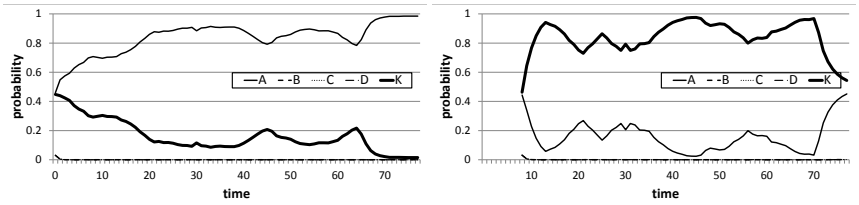
***Case 1*: situation with four people, two authorized (tags A and B), and two intruders - Fig. 5**: this scenario presents several difficulties. Three people, one authorized (ν_1) with tag A and two intruders (ν_3, ν_4), walk in group and after a while (around time 60) the intruders move away and then leave the scene. So, because they are very close to each other, it is not surprising that tag A is mapped to the wrong person for some time. It is worth noting, however, that System is very confident that the tag A and the two intruders are part of the group, because the group localization over time is similar to the one of tag A and there are no other tag with similar localization. The correct mapping is obtained after the group splits. The other authorized person (ν_2) holding tag B, which is located far enough to not be confused with the other tag, is correctly mapped. In this case, Fig. 5(e) also reports the confusion matrix. The poor average precision and recall values are caused by the mapping errors between tags A and $\mathbb{k}$, which are the System's most probable decision at every time. Nevertheless, the errors are bounded to the people in the same group, which cannot be distinguished without other cues.

***Case 2*: three people, two authorized (tags A and B) and one intruder - Fig. 6**: in this case the two authorized people walk side by side, and the intruder follows them at



(a) (b)

Figure 3. Computer graphics rendered images of our scenario: (a) 3D view of the scene, (b) bird-eye view with also locations superimposed. (Courtesy of Davide Baltieri)



(a) Probability of ν_1 (b) Probability of ν_2

Figure 4. *Case 1*, test with one authorized person (ν_1) and one intruder (ν_2). Avg. precision=100.0%, avg. recall=100.0%.

some distance. Aside from the multi-camera issues, this scenario is more complex than the previous one because the two people walk together for the entire test. As a consequence, the mapping process relies only on the similarity between tags and people localization. Once the data allow the system to discover a correct mapping, the reinforcement step excludes this mapping for the other person. The intruder is well mapped because, even if he/she is very often in the same location as the others, he/she moves from one location to another with some delay with respect to them. These differences in the localization are enough to get and keep the correct mapping. By increasing the number of locations, it is possible to cover a larger area, and thus track people for longer time. The more data are available, the more reliable is the mapping.

5. Conclusions

The proposed RFID/camera system shows excellent inference properties in localizing intruders in wide open areas, also in challenging cases where authorized people and intruders follow the same path. Thanks to the sophisticated filtering technique applied to RFID signals and to the evidential fusion architecture based on TBM used as reasoning engine, the noise in the data and the uncertainty in the localization can be successfully handled.

References

[1] D. Arnaud, N. de Freitas, and G. Neil. *Sequential Monte Carlo Methods in Practice*. Springer, 2005.

[2] 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, July 2008.

[3] S. S. Chawathe, V. Krishnamurthy, S. Ramachandran, and S. Sarma. Managing RFID data. In *Proceedings of the Thirtieth international conference on Very large data bases-Volume 30*, pages 1189–1195, 2004.

[4] 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.

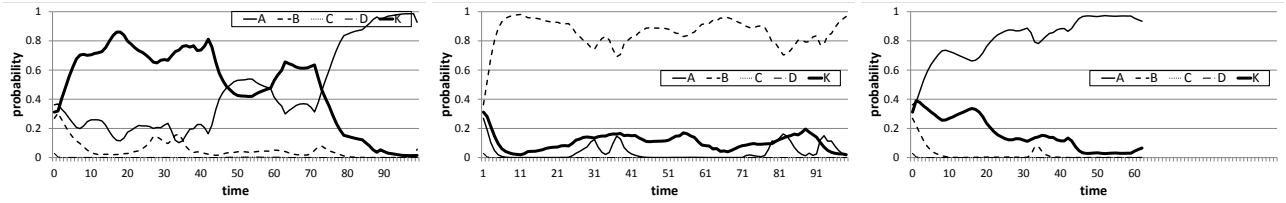
[5] A. Clerentin, L. Delahoche, B. Marhic, M. Delafosse, and B. Allart. An evidential fusion architecture for advanced driver assistance. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 327 –332, Oct. 2009.

[6] R. Cucchiara, M. Fornaciari, A. Prati, and P. Santinelli. Mutual calibration of camera motes and rfids for people localization and identification. In *Proceedings of the ACM/IEEE ICDSC 2010*, Aug. 2010.

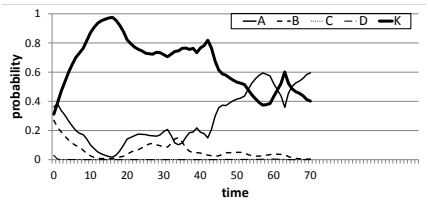
[7] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati. Detecting moving objects, ghosts and shadows in video streams. *IEEE Trans. on PAMI*, 25(10):1337–1342, Oct. 2003.

[8] M. J. Franklin, S. R. Jeffery, S. Krishnamurthy, F. Reiss, S. Rizvi, E. Wu, O. Cooper, A. Edakkunni, and W. Hong. Design considerations for high fan-in systems: The HiFi approach. In *Proc. of the CIDR Conf*, 2005.

[9] M. Guironnet, D. Pellerin, and M. Rombaut. A fusion architecture based on tbm for camera motion classification. *Image Vision Comput.*, 25:1737–1747, November 2007.



(a) Probability of ν_1 (b) Probability of ν_2 (c) Probability of ν_3

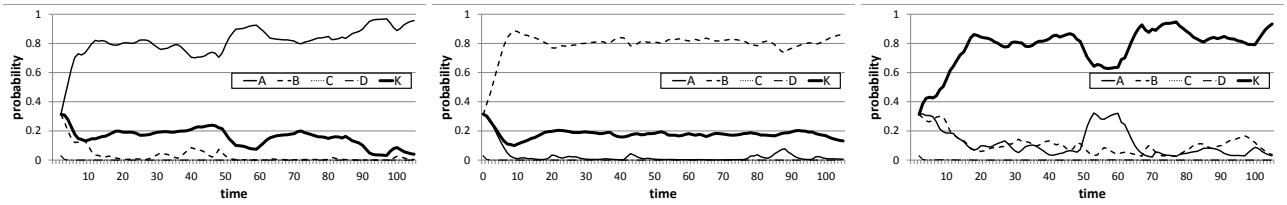


(d) Probability of ν_4

	A	B	C	D	K
A	39	0	0	0	61
B	0	99	0	0	0
C	0	0	0	0	0
D	0	0	0	0	0
K	77	0	0	0	57

(e) Confusion matrix

Figure 5. *Case 1*, test with two authorized people (ν_1 and ν_2) and two intruders (ν_3 and ν_4). Avg. precision=57.6%, avg. recall=56.0%.



(a) Probability of ν_1 (b) Probability of ν_2 (c) Probability of ν_3

Figure 6. *Case 2*, test with two authorized people (ν_1 and ν_2) and one intruder (ν_3). Avg. precision=99.4%, avg. recall=99.4%.

[10] A.-L. Jousselme, D. Grenier, and E. Bosse. A new distance between two bodies of evidence. *Information Fusion*, 2(2):91 – 101, 2001.

[11] B. Kanagal and A. Deshpande. Online filtering, smoothing and probabilistic modeling of streaming data. In *Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on*, pages 1160–1169, 2008.

[12] N. Khoussainova, M. Balazinska, and D. Suciu. Probabilistic event extraction from RFID data. pages 1480–1482, 2008.

[13] 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.

[14] L. R. Rabiner. Readings in speech recognition. chapter A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, pages 267–296. 1990.

[15] J. Rao, S. Doraiswamy, H. Thakkar, and L. S. Colby. A deferred cleansing method for RFID data analytics. In *Proceedings of the 32nd international conference on Very large data bases*, pages 175–186, 2006.

[16] C. Ré, J. Letchner, M. Balazinksa, and D. Suciu. Event queries on correlated probabilistic streams. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, pages 715–728, 2008.

[17] P. Smets. The transferable belief model. *Artificial Intelligence*, 66(2):191–234, 1994.

[18] P. Smets. Analyzing the combination of conflicting belief functions. *Information Fusion*, 8(4):387–412, 2007.

[19] T. Tran, C. Sutton, R. Cocci, Y. Nie, Y. Diao, and P. Shenoy. Probabilistic inference over RFID streams in mobile environments. In *IEEE International Conference on Data Engineering*, pages 1096–1107, 2009.

[20] M. Valera and S. Velastin. Intelligent distributed surveillance systems: a review. *Vision, Image and Signal Processing, IEE Proceedings* -, 152(2):192 – 204, Apr. 2005.

[21] R. Vezzani and R. Cucchiara. Ad-hoc: Appearance driven human tracking with occlusion handling. In *First International Workshop on Tracking Humans for the Evaluation of their Motion in Image Sequences (THEMIS'2008), in conjunction with BMVC 2008*, 2008.

[22] J. Wickramasuriya, M. Datt, S. Mehrotra, and N. Venkatasubramanian. Privacy protecting data collection in media spaces. In *Proceedings of the 12th annual ACM international conference on Multimedia*, MULTIMEDIA '04, pages 48–55, New York, NY, USA, 2004. ACM.

[23] W. Zhang, S.-C. S. Cheung, and M. Chen. Hiding privacy information in video surveillance system. In *Proc. of IEEE Int'l Conference on Image Processing*, pages 868–871, 2005.