Identification of Intruders in Groups of People using Cameras and RFIDs

Rita Cucchiara[1], Michele Fornaciari[1], Razia Haider[1], Federica Mandreoli[1,3], Andrea Prati[2]
[1]DII - University of Modena and Reggio Emilia - Modena, Italy
[2]DiSMI - University of Modena and Reggio Emilia - Reggio Emilia, Italy
[3]IEIIT - BO/CNR, Bologna, Italy

Abstract—**The identification of intruders in groups of people moving in wide open areas represents a challenging scenario where coordination between cameras can be certainly used but this solution is not enough. In this paper, we propose to go beyond pure vision-based approaches by integrating the use of distributed cameras with the RFID technology. To this end, we introduce a system that "maps" RFID tags to people detected by cameras by using sophisticated techniques to filter the singular modalities and an evidential fusion architecture, based on Transferable Belief Model, to combine the two sources of information and manage conflict between them. The conducted experimental evaluation shows very promising results, especially in treating groups of people.**

I. Introduction

The requirements of safety and security in wide open areas call for the use of algorithms for identifying intruders which are not authorized to stay in the area. Considering wide areas with no obliged entrances, standard techniques (such as badges, or fingerprint authentication, face recognition or standard keys) are not the right choice. Besides the need of a specific point of entrance, these techniques also assume a collaborative behavior, which is unlikely for intruders.

In addition to their identification, it is straightforward to understand that localization of intruders is equally important. For instance, it may be useful for acquiring zoomed images of the intruder and proceeding with recognition of gathering of evidences. Actually, these two tasks 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. Coordination between cameras can be used, in order to have PTZ cameras zooming on the person's face while keeping the other camera fixed, but this solution becomes unfeasible in the case of multiple targets to be identified.

This problem is even more challenging when groups of people are present since it is often difficult to distinguish, within a group, authorized people from intruders. Pure vision-based approaches to handle groups of people have been proposed in the past [1], but, generally speaking, 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 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.

With these premises, this paper proposes the joint use of cameras and RFIDS in real noisy and complex wide open areas for intruder localization. The singular modalites are filtered with sophisticated techniques 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. Moreover, an evidential fusion architecture, based on Transferable Belief Model - TBM [2], 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.

II. Related Works

This 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 [3] 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.

978-1-4577-1707-9/11/$26.00 ©2011 IEEE

Another example is provided by the work in [4] where a robot simultaneously interacts with two or more people and has to identify them with a passive-type RFID reader and floor sensors. In [5] 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.

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 assets management, object and people authentication, health care etc. Nevertheless, the data management in these RFID applications poses a number of challenges [6]. 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) [7]. To this end, several techniques have been proposed for the analysis and processing of raw noisy RFID data. Among those, the most effective are the ones that exploit the probabilistic nature of RFID data and manage their inherent uncertainty in the form of probabilities and correlations (see e.g. [7], [8], [9]). 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 [7], [8]. It is worth noting 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). In this regard, the combined RFID/camera system we propose differs from this vision: 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.

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 [10] and for developing a system for advanced driver assistance [11]. The work in [11] 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.

III. System Description

The two sources of information in our system are the current frame f_t provided by the distributed cameras and analyzed by the *Video processing* module, and the RFID signals r_t 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:

- $\mathcal{V}_t$ as the visual objects (typically people) detected by the video processing module at time t, $\mathcal{V}_t = (\nu_{i=1,\dots,n_t})$;
- $\mathcal{T}_t$ as the tags identified by the RFID processing module at time t, $\mathcal{T}_t = (\tau_{i=1,\dots,s_t})$;
- $\mathcal{L}$ 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 $\mathcal{V}_t$ 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 [12] and people are tracked along time using a probabilistic appearance-based tracking algorithm [13] that resulted to be very robust to occlusions. Finally, the tracking among different cameras is obtained using the homography-based approach presented in [1] which also exploits the automatically-learned epipolar geometry of the scene to assign the same label/ID to the same person seen from different cameras.

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 $P\left(L_t^{\tau_j}\right)$, as input and, for each visual object (or people in our case) ν_i and each tag τ_j, it outputs the confidence C of the mapping $\langle\nu_i, \tau_j\rangle$. The higher is $C\langle\nu_i, \tau_j\rangle$ the higher is the confidence that the tag τ_j is held by the person ν_i. The more confident mappings are found, the more precisely intruder can be identified (and then localized) among the people not mapped to any tag.

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

A. RFID Processing

The *RFID Processing* module makes use of a Hidden Markov Model (HMM) [14] in order to infer the values of $L_t^{\tau_i}$, i.e. the variable of interest that is not directly observable, on the basis of the sequence of the observed sensor readings that are made up of four types of information: 1) the TagID the reading is concerning to; 2) the AntennaID the tag is seen by; 3) the timestamp of the reading; 4) the Received Signal Strength Indicator (RSSI) of the reading.

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 includes: 1) the *initial states probability* $P\left(L_0\right)$ of each state, assumed to be uniform distribution; 2) the *state transition probability* $P\left(L_t|L_{t-1}\right)$, i.e. the probabilities of moving from each location at time $t-1$ to each location at time t; 3) the *observation probability* $P\left(R_t|L_t\right)$; in order to learn this CPD, first we learn conjunctive probability $P\left(R_t, L_t\right)$ by using statistical method called *Maximum Likelihood Estimation (MLE)* and then calculate observation probability $P\left(R_t|L_t\right)$ using Bayes' theorem.

Inference is applied through an HMM using an approximate scheme, say *Particle Filtering* [15], that ensures tractability in different real world scenarios. Specifically, the algorithm works by computing and constantly maintaining sets of particles that describe the historical and present states of the model by iteratively executing the following steps:

Initialization: an initial set of particles ($p_0^{i=1,\dots,m}$) is created by random sampling from the initial states probability $P\left(L_0\right)$.

Prediction: for each particle p_{t-1}^i at time $t-1$ a new particle p_t^i is created for time t by sampling from $P\left(L_t|L_{t-1}\right)$.

Filtering: each particle p_t^i is updated by assigning a weight based on the values of the observations at time t and on the given observation probability $P\left(R_t|L_t\right)$. Particles which are close to the observed values will have higher weights as compared to the those which are far from the observed values.

Re-sampling: the particles updated in the *Filtering* step are re-sampled on the basis of the weights assigned.

Broadly speaking, each particle p_t^i represents a guess about the location of tag τ_i. To compute the posterior probability $P\left(L_t^{\tau_i}\right)$ we can indeed simply count the number of particles in each location and divide it by the total number.

B. Reasoning Engine

The reasoning engine has the main objective to fuse the inference coming from video and RFID processing modules by means of the Transferable Belief Model (TBM) [2].

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* (FoD) $\Delta = \{H_1, \dots, H_b\}$ containing b mutually and exhaustive hypotheses 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. Every subset $A \subseteq \Delta$ where $m(A) > 0$ is called *focal element*. Work with focal elements only avoids the exponential complexity of the TBM. The symbol $|\cdot|$ indicates the cardinality of the set.

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 ($m(H_i) = 1$). It is a powerful model to deal with *uncertainty*, which may results from sensor noise, misreading or semantic noise.

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

a) FoD updating: At time t, we define the FoD $\mathcal{V}_t$ as the set of people seen and the FoD $\mathcal{T}_t$ as the set of tags sensed.

Given that a new person $\nu_i \notin \mathcal{V}_t$ appears at time $t+1$, the hypothesis ν_i is added to FoD: $\mathcal{V}_{t+1} = \mathcal{V}_t \cup \{\nu_i\}$. To reflect this change in all the belief functions $bMap^{\nu_j}$ (see *e)* in the following) we need to apply a *deconditioning* process [16], which simply appends (in the set union sense) to each focal element the missing hypothesis. If at time $t+1$ a person $\nu_i \in \mathcal{V}_{t-1}$ disappear, the hypothesis ν_i is removed from FoD: $\mathcal{V}_{t+1} = \mathcal{V}_t \backslash \{\nu_i\}$, and ν_i is removed from the focal elements of all $bMap^{\nu_j}$ in the *conditioning* process. Moreover, a normalization step is obtained by transferring the mass of the empty set to the total ignorance set ($m(\mathcal{V}_{t+1}) = m(\mathcal{V}_t) + m(\emptyset)$).

The same considerations apply when a tag τ_j appears or disappears, except that the FoD is $\mathcal{T}_t$ and the belief functions to modify are $bMap^{\tau_i}$.

b) Belief on locations: The probabilities over the locations of tags $\tau_i \in \mathcal{T}_t$, $P\left(L_t^{\tau_i}\right)$, and visual objects $\nu_i \in \mathcal{V}_t$ (or people, in our case), $P\left(L_t^{\nu_i}\right)$, are translated in a Bayesian belief function [17] on the FoD $\Delta \equiv \mathcal{L}$, 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: $m_t(\lambda_j) = P\left(L_t^{\tau_i} = \lambda_j\right), \forall\lambda_j \in \mathcal{L}$, 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 [17], we use *Dubois-Prade's conjunctive combination rule* because it merges coherent information in a conjunctive way and conflicting ones in a disjunctive way:

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

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

c) 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 mapping $\langle\nu_i, \tau_j\rangle$ between ν_i and τ_j derived from all information available at this moment. Defining $fe(x)$ as the set of focal elements 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 [18]:

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

d) Evidence generation: The above mentioned similarity values are exploited to generate new pieces of information (*evidences*), encoded as a *bba*, representing all the knowledge that System is able to extract from data available at time t.

First, for each person $\nu_i \in \mathcal{V}_t$, the *bba* on the FoD $\mathcal{T}_t$ encodes the belief on which tag(s) can be held by ν_i. Let $\Psi(\nu_i) = \{\psi(\nu_i, \tau_1), \ldots, \psi(\nu_i, \tau_r)\}$ be the set of similarity values between ν_i and the $r = |\mathcal{T}_t|$ tags sensed at this moment ordered by decreasing value. We create the focal elements of the evidence using the following criterion, where $\Gamma_j = \bigcup_{1\leq i\leq j,\;\; j\leq r} \tau_i$:

$$m_t^{\nu_i}(\Gamma_j) = \begin{cases} \psi(\nu_i, \tau_{j-1}) - \psi(\nu_i, \tau_j) & \text{, if } j < r \\ \psi(\nu_i, \tau_j) & \text{, if } j = r \end{cases} \tag{3}$$

$$m_t^{\nu_i}(\mathcal{T}_t) = (1 - \psi(\nu_i, \tau_1)) \tag{4}$$

Second, for each tag $\tau_j \in \mathcal{T}_t$, the *bba* on the FoD $\mathcal{V}_t$ encodes the belief on which person(s) can hold the tag τ_j. Let $\Psi(\tau_j) = \{\psi(\nu_1, \tau_j), \ldots, \psi(\nu_q, \tau_j)\}$ be the set of similarity values between τ_j and the $q = |\mathcal{V}_t|$ traces seen at this moment ordered by decreasing value. We create the *bba* as before, replacing in (3) and (4) $r, \nu, \tau, \mathcal{T}_t$ with $q, \tau, \nu, \mathcal{V}_t$, respectively. Note that $\psi(\nu_i, \tau_j) \equiv \psi(\tau_j, \nu_i)$ by definition.

e) Belief updating: System entertains a belief on two kinds of mapping. The first is the mapping between a person ν_i and the tags, encoded as a belief function $bMap^{\nu_i}$ on the FoD $\mathcal{T}_t$. For each person $\nu_i \in \mathcal{V}_t$ we combine the relative evidence $m_t^{\nu_i}$ with $bMap^{\nu_i}$. Similarly, the second mapping is between a tag τ_j and the people, encoded as a belief function $bMap^{\tau_i}$ on the FoD $\mathcal{V}_t$. For each tag $\tau_j \in \mathcal{T}_t$ we combine the relative evidence $m_t^{\tau_j}$ with $bMap^{\tau_j}$.

To combine new evidences with System's belief we use eq. (1) with $m_1 = bMap^{\nu_i}$ (resp. $bMap^{\tau_j}$), $m_2 = m_t^{\nu_i}$ (resp. $m_t^{\tau_i}$) and $A, W, X, Y, Z \subseteq \mathcal{V}_t$ (resp. $\mathcal{T}_t$).

f) Betting: The belief function $bMap^{\nu_i}$ represents the System knowledge updated with all information available till now on which tag(s) can be held by the person ν_i.

For each $bMap^{\nu_i}$ System constructs a probability function Bet^{ν_i}, over the set of hypotheses $\mathcal{T}_t$, in order to make the optimal decision. If System must decide *now* what is the right tag to map to ν_i, the decision is taken accordingly to Bet^{ν_i}: this approach translates the saying that "beliefs guide our action". We then apply the following *pignistic transformation*, where $Bet^{\nu_i}(\tau_j)$ denotes the probability that the tag τ_j is held by the person ν_i:

$$Bet^{\nu_i}(\tau_j) = \sum_{A:\tau_j \in A \subseteq \mathcal{T}_t} \frac{m(A)}{|A|(1 - m(\emptyset))} \tag{5}$$

The same considerations are valid also when System constructs a probability function Bet^{τ_j} over the set of hypotheses $\mathcal{V}_t$ starting from $bMap^{\tau_j}$ to find the right person to map with τ_j.

g) Decision: System estimates, at best of its knowledge, which mappings $\langle \nu_i, \tau_j \rangle$, among all possible mappings between a single person and a single tag, are the most likely. In the set $fe(bMap^{\nu_i})$, the focal element Q^{ν_i} with highest mass is, thus, the (set of) hypothesis which has the highest support. Similarly, in the set $fe(bMap^{\tau_j})$, the focal element R^{τ_j} with highest mass is the (set of) person(s) which is more likely to hold τ_j.

The confidence C on $\langle \nu_i, \tau_j \rangle$ is determined as follows:

$$C\langle \nu_i, \tau_j \rangle = \frac{1}{2}\left(\frac{m(Q^{\nu_i})}{|Q^{\nu_i}|} Bet^{\tau_j}(\nu_i) + \frac{m(R^{\tau_j})}{|R^{\tau_j}|} Bet^{\nu_i}(\tau_j) \right) \tag{6}$$

which merges System's knowledge on which tag can be held by a given person with the one on which person can hold a given tag.

By thresholding the confidence on the mappings (see Sec. IV), System is able to determine which people in the scene are authorized, also assigning them their own tag. The more high confidence mappings are discovered, the more confident is System to detect as intruder who is not mapped to any tag.

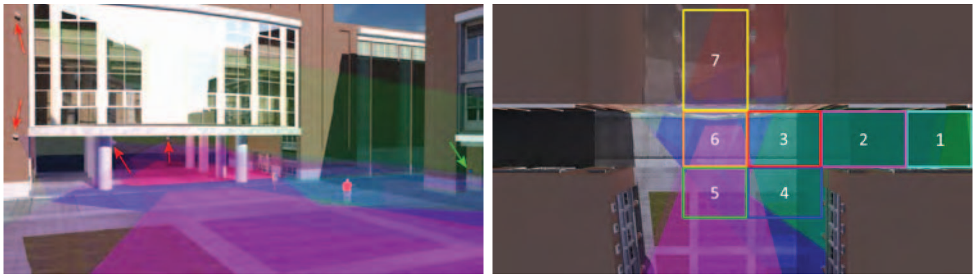
C. On the choice of locations

The ultimate goal of the system is to identify and locate possible intruders while locations are the mean which data coming from both the camera and RFID processing modules are mapped to. Therefore, in order to maximize the cooperation between the two modules the open area can be divided 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 high to allow a substantial amount of location changes to be identified by the RFID and video modules. 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 [19]).

IV. Experimental Results

For evaluating the effectiveness of our system 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, especially when dealing with groups of people.

Fig. 1 shows the overview of the testbed: Fig. 1(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. 1(b) reports a bird-eye view of the setup with the locations represented by bounded areas and the antenna indicated by a green arrow. Four cameras and seven locations have been used, allowing a larger scene coverage (thanks to the consistent labeling techniques described in Section III) and considering a more realistic scenario. Upon this challenging setup, we have collected data from cameras and RFID tags in several experiments.



(a) (b)

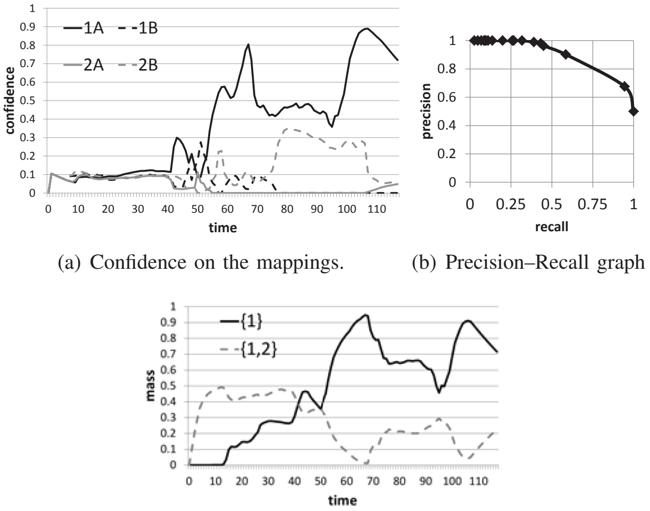
Fig. 1. Computer graphics rendered images: (a) 3D view of the scene, (b) bird-eye view with also locations superimposed (by Davide Baltieri).

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.

***Test 1*: two people, two authorized (tags τ_A and τ_B) (Fig. 2)**. The two authorized people ν_1 and ν_2, holding respectively τ_A and τ_B, walk side by side for a while and then split (around time 50). System is able to manage the group of people thanks to the nature of TBM. As long as the group moves together it is impossible to discover the correct mappings: System believes that every combination of people and tags could be a possible mapping. The data available after the splitting allow to refine System's knowledge, *transferring the belief* to more specific sets of hypothesis (Fig. 2(c), where around time 50 the belief is transferred from the set $\{1, 2\}$ to the correct set $\{1\}$) until a mapping is found. We also compute the average precision and recall by changing the threshold applied on the confidence of the mappings reported in Fig. 2(a). The different values are collected in the precision–recall graph shown in Fig. 2(b).

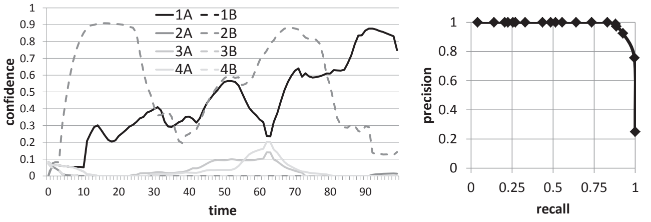
***Test 2*: four people, two authorized (tags τ_A and τ_B), and two intruders (Fig. 3)**. The authorized person ν_1 holding tag τ_A walks nearby a group of two intruders. Because of the short distance between ν_1 and the intruders, they are always detected in the same location, except when they cross an edge between two locations (around time 10). The history coded in the belief functions allows to keep the correct mapping $\langle \nu_1, \tau_A \rangle$ from that moment on. The two intruders ν_3 and ν_4 are correctly found as not holding any tag. The other authorized person (ν_2) holding tag τ_B, which is located far enough to not be confused with the others, is correctly mapped $\langle \nu_2, \tau_B \rangle$ for most of the time.



(a) Confidence on the mappings. (b) Precision–Recall graph

(c) Some focal elements of $bMap^{\tau_A}$; $1 \equiv \nu_1$ and $2 \equiv \nu_2$

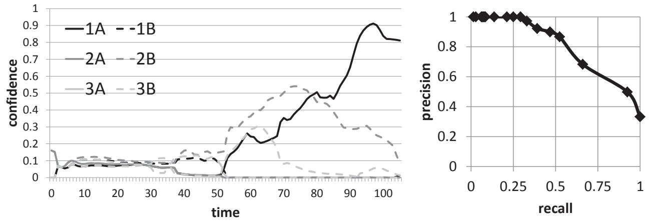
Fig. 2. Test 1: The correct mappings are: $\langle \nu_1, \tau_A \rangle \equiv$ 1A and $\langle \nu_2, \tau_B \rangle \equiv$ 2B.



(a) Confidence on the mappings (b) Precision–Recall graph

Fig. 3. Test 2: The correct mappings are: $\langle \nu_1, \tau_A \rangle \equiv$ 1A and $\langle \nu_2, \tau_B \rangle \equiv$ 2B.

***Test 3*: three people, two authorized (tags τ_A and τ_B), and one intruder (Fig. 4)**. The two authorized people ν_1 and ν_2 walk side by side, while the intruder ν_3 follows them at some distance. System is able to manage the ambiguous situation keeping the confidences on the mappings very low when ν_1 and ν_2 are very close and their localization is the same. When the two people split (around time 65), however, data on localization allow to recognize the correct mappings and their confidences increase quickly. The confidence on the mappings for the intruder ν_3 is correctly always very low, because his position is different from the one of both tags. The precision and recall values are lower than test 1 because of the long-lasting side-by-side walk of ν_1 and ν_2.

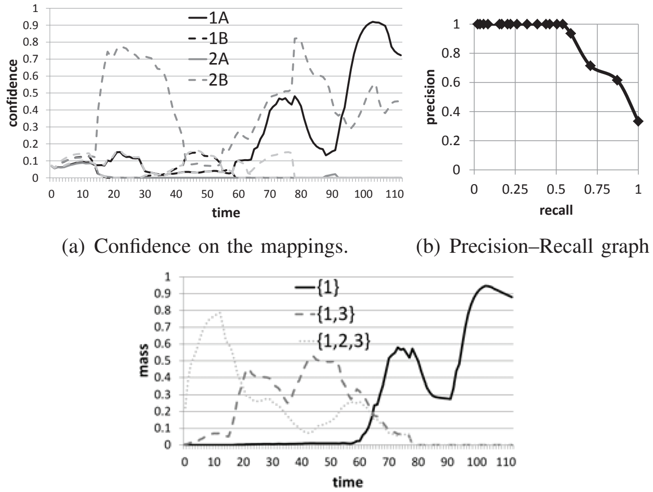


(a) Confidence on the mappings.

(b) Precision–Recall graph

Fig. 4. Test 3: The correct mappings are: $\langle \nu_1, \tau_A \rangle \equiv$ 1A and $\langle \nu_2, \tau_B \rangle \equiv$ 2B.

***Test 4*: three people, two authorized (tags τ_A and τ_B), and one intruder (Fig. 5)**. The authorized person ν_1 (tag τ_A) and the intruder ν_3 walk side by side, and the other authorized person ν_2 follows them at some distance. The correct mapping $\langle \nu_2, \tau_B \rangle$ is soon found, while $\langle \nu_1, \tau_A \rangle$ has low confidence until ν_1 and the intruder split (around time 60). Analyzing some focal elements of the belief function $bMap^{\tau_A}$ (Fig. 5(c)), is clearly visible how System first allocates most of the mass to the broader set of hypothesis (i.e. System believes that τ_A can be held by every person) and progressively transfers the mass to narrower sets, until only one hypothesis (ν_1) remains.



(a) Confidence on the mappings.

(b) Precision–Recall graph

(c) Some focal elements of $bMap^{\tau_A}$; $1 \equiv \nu_1$, $2 \equiv \nu_2$, $3 \equiv \nu_3$

Fig. 5. Test 4: The correct mappings are: $\langle \nu_1, \tau_A \rangle \equiv$ 1A and $\langle \nu_2, \tau_B \rangle \equiv$ 2B.

***Degeneration of the system*:** In some particular conditions, e.g. when an intruder ν_a is present in the scene, and, simultaneously, a tag τ_j is sensed but the relative authorized person is not detected, System is forced to map that tag with the intruder. The similarity values in III-B.c would be very low, and the evidence generated in III-B.d would assign correctly most of the mass to the $\mathcal{T}_t$ set. Nevertheless, because $\mathcal{T}_t$ contains τ_j only, the resulting mapping is $\langle \nu_a, \tau_j \rangle$, which is incorrect.

To avoid these wrong mappings, at every t a dummy person ν_x and a dummy tag τ_k (on which System has no belief on mapping) are added to *close the world* on the sets of hypotheses. Both of them are localized with a uniform distribution over the set of locations $\mathcal{L}$. Thus, $\mathcal{V}_t$ and $\mathcal{T}_t$ always contain, respectively, the identifiers ν_x and τ_k.

V. 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 in group. 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] S. Calderara, A. Prati, and R. Cucchiara, “Hecol: Homography and epipolar-based consistent labeling for outdoor park surveillance,” *Computer Vision and Image Understanding*, vol. 111, no. 1, pp. 21–42, 2008.

[2] P. Smets, “The transferable belief model,” *Artificial Intelligence*, vol. 66, no. 2, pp. 191–234, 1994.

[3] 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*, 2005, pp. 868–871.

[4] 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*, pp. 545 –552, aug. 2008.

[5] 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*. Washington, DC, USA: IEEE Computer Society, 2009, pp. 553–558.

[6] 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*, 2004, pp. 1189–1195.

[7] C. Ré, J. Letchner, M. Balazinksa, and D. Suciu, “Event queries on correlated probabilistic streams,” in *Proc. of the ACM SIGMOD international conference on Management of data*, 2008, pp. 715–728.

[8] N. Khoussainova, M. Balazinska, and D. Suciu, “Probabilistic event extraction from RFID data,” 2008, pp. 1480–1482.

[9] 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*, 2008, pp. 1160–1169.

[10] M. Guironnet, D. Pellerin, and M. Rombaut, “A fusion architecture based on tbm for camera motion classification,” *Image Vision Comput.*, vol. 25, pp. 1737–1747, November 2007. [Online]. Available: http://portal.acm.org/citation.cfm?id=1280281.1280314

[11] 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*, Oct. 2009, pp. 327 –332.

[12] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, “Detecting moving objects, ghosts and shadows in video streams,” *IEEE Trans. on PAMI*, vol. 25, no. 10, pp. 1337–1342, Oct. 2003.

[13] R. Vezzani and R. Cucchiara, “Ad-hoc: Appearance driven human tracking with occlusion handling,” in *Intl WS on Tracking Humans for the Evaluation of their Motion in Image Sequences (THEMIS)*, 2008.

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

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

[16] F. Janez and A. Appriou, “Theory of evidence and non-exhaustive frames of discernment: Plausibilities correction methods,” *International Journal of Approximate Reasoning*, vol. 18, no. 1-2, pp. 1 – 19, 1998.

[17] P. Smets, “Analyzing the combination of conflicting belief functions,” *Information Fusion*, vol. 8, no. 4, pp. 387–412, 2007.

[18] A.-L. Jousselme, D. Grenier, and E. Bosse, “A new distance between two bodies of evidence,” *Information Fusion*, vol. 2, no. 2, pp. 91 – 101, 2001.

[19] 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.