An Evidential Fusion Architecture for People Surveillance in Wide Open Areas

M. Fornaciari[1], D. Sottara[2], A. Prati[3], P. Mello[2], and R. Cucchiara[1]

[1] DII - Univ. of Modena and Reggio E. - Via Vignolese, 905 - 41125 Modena, Italy
[2] DEIS - Univ. of Bologna - Via Zamboni, 33 - 40126 Bologna, Italy
[3] DISMI - Univ. of Modena and Reggio E. - Via Amendola, 2 - 42122 R.E., Italy

Abstract. A new evidential fusion architecture is proposed to build an hybrid artificial intelligent system for people surveillance in wide open areas. Authorized people and intruders are identified and localized thanks to the joint employment of cameras and RFID tags. Complex Event Processing and Transferable Belief Model are exploited for handling noisy data and uncertainty propagation. Experimental results on complex synthetic scenarios demonstrate the accuracy of the proposed solution.

1 Introduction and Related Works

There are several fields in Information Technologies research where specific strategies are required for *localizing* and *identifying* objects, such as automatic logistics of goods, surveillance, industrial pick-and-place applications, etcetera. These 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 (and look at more objects simultaneously), whereas identification often requires a close-up of the object. This is particularly true in video surveillance and when the objects are people: identification might require zooming on the person's face and localization needs an unzoomed view to find correct position with respect to the scene, even if many people are potentially present.

Instead of cameras, alternative sensors can be used for localization and identification purposes. Among the many, RFID (Radio Frequency IDentification) tags gained much attention thanks to their ease of use, low cost and touchless way-of-reading. The identification with RFID tags is accurate, but there is a severe limitation: true positives are detected, i.e. people wearing a tag, but not true negatives, i.e. people without the tag. Regarding localization, there have been previous attempts [1, 2] which use multiple tags or readers to assess people's locations using triangulation and Received Signal Strength Indicator (RSSI) value. These approaches, however, do not guarantee a sufficient accuracy.

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. One may think to the application of this technology to construction working sites, where wide open areas with no obliged entrances are considered. In this scenario, illumination changes, occlusions and reflexes can make hard the task for computer vision algorithms applied to cameras, while multiple signal sources and the presence of metallic objects can introduce much noise in RFID signals.

With these premises, this work proposes to jointly use cameras and RFIDs and to take 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. The envisaged application (localization and identification of people in wide open areas) presents two strict requirements: (i) to provide a quick response, the system must work with incrementally acquired observations, so the knowledge must be updated online; (ii) both sensory modalities (cameras and RFIDs) are very heterogeneous and present a high degree of noise and uncertainty in the measurements. With these requirements, we propose a system which includes a reasoning engine with a Complex Event Processing module capable to handle uncertainty and an evidential fusion architecture (based on Transferable Belief Model - TBM [3]) to process imprecise (or missing) data, to combine various sources of information and to manage the conflict between the sources. It is an excellent tool for filtering false alarms by an optimal management of the uncertainty estimation.

TBM has been used in the literature for different applications, such as for the classification of the camera motion [4] and for developing a system for advanced driver assistance [5], but at the best of our knowledge this is the first case of application to people surveillance. 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.

The final proposal is defined as a hybrid artificial intelligence system (HAIS) [6, 7], which represents an useful tool since they combine both symbolic and subsymbolic paradigms for increasing the robustness in problem-solving problems [8]. These systems are becoming more and more popular due to their ability to handle aspects such as imprecision, uncertainty or high dimensionality of data. In recent past, they have been used for computer network security and intrusion detection [8], for home care assistance through a multi-sensory architecture [9], or for minimizing the energy consumption in heating systems [10].

2 The Evidential Fusion Architecture

The surveillance system processes data coming from both cameras and RFID tags in order to generate a correct mapping between people and tags. In this way the system is able to authenticate *authorized people* (which are provided with one – and only one – tag) and to recognize *intruders* (which have no tag) and even *unfair people* (which hold more than one tag). This latter case, which can be potentially treated with the proposed evidential fusion architecture, will not be considered in this paper. The final system will process real-time data coming from real cameras and RFIDs (a first example is reported in [11]). This paper, however, will focus mainly on the reasoning engine and will test our novel proposal on synthetic yet realistic data only.

The target scenario is a wide outdoor area surveilled by a certain number of cameras with given fields of view (FoV) and RFID antennas. In addition, several locations are defined in this area in order to have the same RSSI behavior within a location and different behavior when changing the location. This can be done manually or automatically, for instance following the approach described in [11].

2.1 (Uncertain) Complex Event Processing

Each piece of information coming from the sensors can be considered as an *event*: to handle them properly, we adhere to the principles of Complex Event Processing [12] (CEP), an emerging approach to model and implement event-oriented systems. In our CEP, an *Event* is a significant state change in the observed environment at a given time, encoded using a symbolic representation which, in addition to the pieces of information required to formalize the state change, explicitly keeps a notion of the instant the event took place. Events assume importance due to their temporal, causal and hierarchical correlations with other events. The latter, in particular, allows to define a complex event A by aggregation of other events $E_{i:1..n}$ defined at a lower level of abstraction. Primitive events, then, are generated by interfaces with the environment, such as cameras and RFIDs in our case, while complex ones are raised internally. The event handlers are usually defined in a symbolic way, for example using reactive rules [13]. In CEP, however, events are usually defined and known with precision, while in our system the sensors provide only probability distributions for each person and tag in every location. An uncertain event, then, is a pair $< A, \pi >$ associating a traditional event descriptor A to an uncertainty model, such as a probability degree π. When a complex event is generated by aggregation, its uncertainty degree is obtained by analysis and combination of the uncertainties associated to the component events: $< A, \pi_A > \doteq < f(E_{j:1..n}), g(E_{j:1..n}, \pi_{j:1..n}) >$. For the scope of this paper, f denotes any reactive rule used to aggregate events. The uncertainty models and the combination rules g are described in section 3.

Given a set of persons $\Pi = \{P_{i:1..\pi}\}$, a set of tags $\Theta = \{T_{j:1..\tau}\}$ and a set of locations $\Lambda = \{L_{k:1..\lambda}\}$, the relevant events in the proposed surveillance systems are as follows. All the events are instantaneous and marked with a timestamp t.

Person in Location $personInLoc(P_i, L_k, t)$: P_i was located in L_k. This primitive event is generated for each frame by a *Video Module* analyzing the scene derived from the camera with frame rate ν_C.

Tag in Location $tagInLoc(T_j, L_k, t)$: T_j was located by in L_k. Each tag transmits its identification code with frequency ν_T: the antenna provides also a RSSI value that can be analyzed by a *RFID Module* using a Hidden Markov Model to determine the position of the tag.

Person Movement P_i moved from L_k to L_h:
$personInLoc(P_i, L_k, t-1) \wedge personInLoc(P_i, L_h, t) \Rightarrow mov(P_i, L_{kh}, t)$

Tag Movement T_j moved from L_k to L_h:
$tagInLoc(T_j, L_k, t-1) \wedge tagInLoc(T_j, L_h, t) \Rightarrow mov(T_j, L_{kh}, t)$

Movement Correlation P_i and T_j performed a similar movement:
$mov(P_i, L_{kh}, t) \wedge mov(T_j, L_{kh}, t) \Rightarrow corr(P_i, T_j, t)$

Tag Owner P_i holds T_j: $corr(P_i, T_j, t) \Rightarrow holds(P_i, T_j, t)$
Intruder P_i was known not to hold any valid tag:
$\exists P_i : \forall T : \neg holds(P_i, T, t) \Rightarrow alarm(P_i, t)$

3 Implementation

The probabilistic nature of primitive events $XinLoc()$ (where "X" stands for either "person" or "tag") is not the only source of uncertainty, since the chosen aggregation rules are not certain themselves. For example, the fact that a person P_i and a tag T_j perform the same movement at the same time may not be sufficient to decide that P_i holds T_j, possibly because there is more than one person and/or tag which performed the same movement, or the location of P_i (resp T_j) could not be determined with precision. On the other hand, if a person and a tag continue moving in the same fashion over time, the mapping between the two may become stronger. To model such concepts, we chose the TBM [3] because (i) it supports the combination of different pieces of evidence coming from different sources and (ii) it allows the information to be updated in time, thus permitting to propagate the uncertainty associated to the events.

Each probability function provided by Video and RFID modules induces a set of isopignistic belief functions on the frame of discernment Λ [14], of which we keep the q-least committed one. The movement $mov(X, L_{kh}, t)$ of a person/tag X from location L_k to L_h at time t is *possible* only when it happens between *adjacent* locations, i.e. start and end locations are reachable without passing through any other location. For each person/tag, belief functions over Λ at time $t-1$ and t are combined using the following combination rule:

$$m^M(W) = \sum_{Y,Z \subseteq \Lambda : W = (\{Y \times Z\} \cap M) \subseteq \Lambda \times \Lambda} m^{\Lambda}_{t-1}(Y) m^{\Lambda}_{t}(Z) \, . \tag{1}$$

In other words, we obtain a new belief function on the power set 2^M over the frame of discernment of possible movements $M = \{L_{kh} \equiv (L_k, L_h) : L_k, L_h \in \Lambda, L_k \text{ is adjacent to } L_h\} \subseteq \Lambda \times \Lambda$.

The *correlation* between P_i and T_j is the support to the fact that P_i holds T_j. It can be derived comparing belief functions on movements of people and tags with a similarity measure which accounts for the similarity between focal elements through the Jaccard index [15], where F_i is the set of focal elements:

$$\psi(F_1, F_2) = \sum_{A \in F_1} \sum_{B \in F_2} m_1(A) \cdot m_2(B) \cdot \frac{|A \cap B|}{|A \cup B|} \, . \tag{2}$$

Then it can be translated, as a basic mass assignment, into the piece of evidence:

$$corr(P_i, T_j, t) \Rightarrow evd^{\Omega}_{P_i} = \begin{cases} m(T_j) = \psi(F_{P_i}, F_{T_j}) \\ m(I \setminus T_j) = 1 - \psi(F_{P_i}, F_{T_j}) \end{cases} \quad , \tag{3}$$

with $\Omega = \Theta \cup \{\Bbbk\}$, where $\Bbbk$ is the identifier of the dummy tag held by an intruder used to *close the world* on the frame of discernment of identifiers. A common sense rule is that people *seen* at time t can hold tags *sensed* at time t or $\Bbbk$ only, and so, for each person, $\psi(F_{P_i}, F_{T_j}) = 0$ for each tag *not sensed* at time t.

System (which is the agent that entertains the belief on people's tags ownership) keeps a belief function for each person (present in the scene) over the frame of discernment Ω representing System's belief that P_i holds a subset of tags in Ω. Among different conjunctive combination rules [16], we use Dubois-Prade's rule to combine the evidence $evd^{\Omega}_{P_i}$ with the belief function of P_i. The use of a normalized combination rule ($m(\emptyset) = 0$) is consistent with the closed world assumption, and also allows to exclude the Smets' rule. The redistribution of conflicting masses over all focal elements of Dempster–Shafer's rule has poor significance in our application.Yager's rule is more cautious but less committed than Dubois–Prade's rule, in which masses resulting from pairs of conflictual focal elements are transferred to the union of these subsets. Moreover, true but conflicting information are combined as a disjunctive combination rule and thus this is the only rule which accounts for a person holding more than one tag allowing to handle the unfair people case mentioned above.

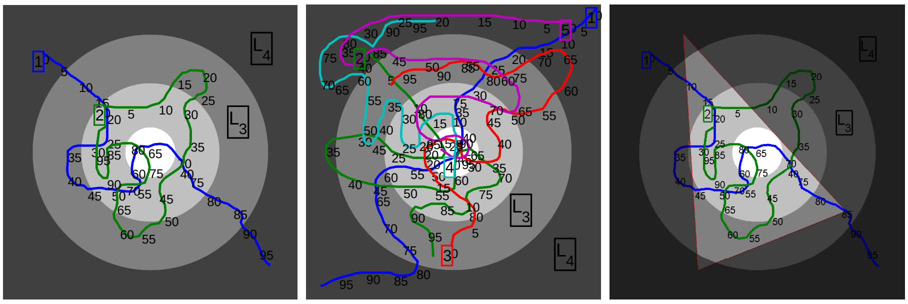
The pignistic transformation on System's beliefs about tags ownership generates a probabilistic distribution over the betting frame Ω which (picking the maximum value) indicates which tag is held by each person. If the maximum value corresponds to $\Bbbk$ then the person is an intruder and an alarm ($alarm(P_i, t)$) must be triggered, otherwise authorized people holding one tag only may be authenticated ($holds(P_i, T_j, t)$). Besides, unfair people may be recognized with a more sophisticated rule which exploits also System's belief over sets of tags.

4 Experimental Results

For evaluating the effectiveness of the proposal we present here a set of synthetic experiments, where the inputs from cameras and RFIDs are realistically simulated. The envisaged layout is composed of locations concentric with the only RFID antenna positioned exactly in the middle of the square area representing the FoV, eventually obtained merging the FoV of multiple cameras (Fig. 1(a,b)).

The radius of each location is determined by the behavior of RSSI over distance from antenna obtained experimentally. This behavior is reproduced first defining the RSSI value as a function of the distance from antenna and then perturbing with white noise with standard deviation σ. Trajectories have been traced manually and so exact position, location and distance from antenna are known. Both ν_C and ν_T are 1 Hz.

While in real applications we intend to use state-of-the-art computer vision techniques, in this simulated context $personInLoc(P_i, L_h, t)$ is α if P_i is in location L_h and $(1-\alpha)/(\lambda-1)$ otherwise, with $\alpha = 0.8$. A Hidden Markov Model with locations as hidden states and RSSI values as observations is trained simulating RSSI values of a tag following a simulated probing trajectory. For tag T_j, $tagInLoc(T_j, L_h, t)$ is the probability of the hidden state L_h given as observation the simulated RSSI value after the application of a median filter with a window size of 3 to further reduce the noise.



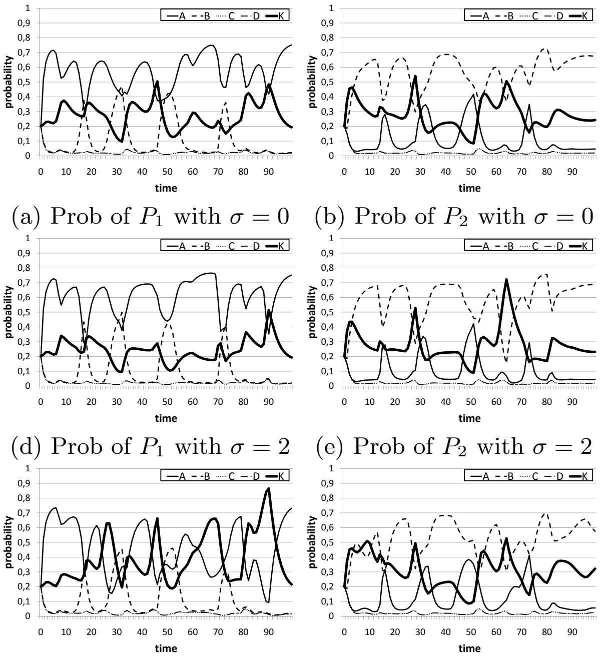
(a) (b) (c)

Fig. 1. Three scenarios considered in our tests. Numbers represent time.

Hereafter three scenarios of increasing level of complexity are proposed. For each person present in the scene, the probability that he/she holds a certain tag is depicted in the graphs of Figs. 2-3. The maximum value at a given time is the System's bet over the mapping between people and tags. The first scenario (Fig. 1(a)) represents two people moving in the scene holding tag A and B respectively. The probability of the correct mapping is always the highest (Fig. 2(a-c)), except for few samples when persons change location: since tags localization is less responsive than people localization due to the HMM, for a short interval the tag and person holding it are, wrongly, localized in different locations. This scenario is analyzed also considering different amount of noise in the RSSI of the tags (Fig. 2(d-i)). Average precision and recall values (see caption of Fig. 2) are very high for $\sigma = 0$ and 2, while recall decreases to 70% in the case of high noise ($\sigma = 5$) which affects localization accuracy. The second scenario (Fig. 1(b)) is the most complex and the precision and recall drop to around 70-80% (Fig. 3). Finally, the third scenario (Fig. 1(c)) only differs from the first one for the FoV which does not cover the whole layout, thus simulating a more realistic scenario. The results in Fig. 4, where the darkened areas represent the intervals in which people are outside the FoV, with avg. precision$=$88.3% and avg. recall$=$77.0%, are comparable with the ones in Fig. 2, showing the robustness of the proposed approach also in more realistic cases. In general, most of the errors are due the assignment of a tag to one of the intruders, mainly because for some time an intruder is in the same location of an authorized person. However authorized people are correctly identified most of the time.

5 Conclusions

The proposed architecture successfully demonstrates to be very accurate on synthetic yet complex data. Though more evidence must be collected on real data, we firmly believe that the proposed methodology, which is conceived to deal with noise and uncertainty typical of real data, will be effective also in that case.



(a) Prob of P_1 with $\sigma = 0$ (b) Prob of P_2 with $\sigma = 0$

(d) Prob of P_1 with $\sigma = 2$ (e) Prob of P_2 with $\sigma = 2$

(g) Prob of P_1 with $\sigma = 5$ (h) Prob of P_2 with $\sigma = 5$

	A	B	C	D	K
A	90	5	0	0	5
B	1	88	0	0	11
C	0	0	0	0	0
D	0	0	0	0	0
K	0	0	0	0	0

(c) CM with $\sigma = 0$

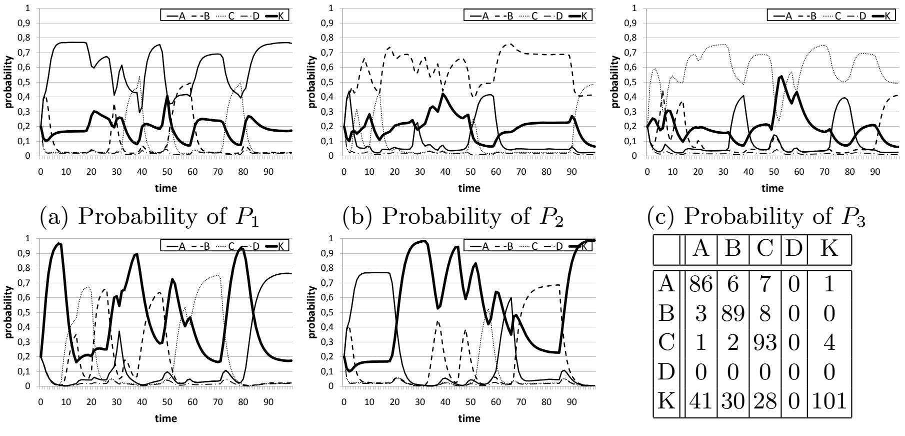
	A	B	C	D	K
A	94	4	0	0	2
B	1	89	0	0	10
C	0	0	0	0	0
D	0	0	0	0	0
K	0	0	0	0	0

(f) CM with $\sigma = 2$

	A	B	C	D	K
A	59	8	0	0	33
B	1	81	0	0	18
C	0	0	0	0	0
D	0	0	0	0	0
K	0	0	0	0	0

(i) CM with $\sigma = 5$

Fig. 2. Scenario with two people P_1 and P_2 holding tag A and B, respectively. Results at different values of σ: $\sigma = 0$ avg prec=96.8%, avg recall=89.0%; $\sigma = 2$ avg prec=97.3%, avg recall=91.5%; $\sigma = 5$ avg prec=94.7%, avg recall=70.0%



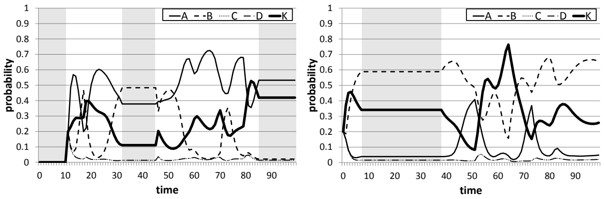
(a) Probability of P_1 (b) Probability of P_2 (c) Probability of P_3

(d) Probability of P_4 (e) Probability of P_5

	A	B	C	D	K
A	86	6	7	0	1
B	3	89	8	0	0
C	1	2	93	0	4
D	0	0	0	0	0
K	41	30	28	0	101

(f) Confusion matrix

Fig. 3. Complex scenario with five people: P_1, P_2 and P_3 hold a tag (A,B and C), while P_4 and P_5 are intruders (K) ($\sigma = 2$): avg prec=74.8%, avg recall=79.6%



(a) Probability of P_1 (a) Probability of P_2

	A	B	C	D	K
A	73	23	0	0	4
B	1	81	0	0	18
C	0	0	0	0	0
D	0	0	0	0	0
K	0	0	0	0	0

(b) Confusion matrix

Fig. 4. Same scenario as Fig. 2, but with different FoV.

References

1. Sanpechuda, T., Kovavisaruch, L.: A review of RFID localization: Applications and techniques. In: ECTI-CON, pp. 769–772 (2008)
2. Xin, H., Janaswamy, R., Ganz, A.: Scout: Outdoor localization using active RFID technology. In: BROADNETS, pp. 1–10 (2006)
3. Smets, P.: The transferable belief model. Artif. Intell. **66**, 191–234 (1994)
4. Guironnet, M., Pellerin, D., Rombaut, M.: A fusion architecture based on tbm for camera motion classification. Image Vision Comput. **25**, 1737–1747 (2007)
5. Clerentin, A., Delahoche, L., Marhic, B., Delafosse, M., Allart, B.: An evidential fusion architecture for advanced driver assistance. In: IEEE/RSJ IROS, pp. 327–332 (2009)
6. Medsker, L.: Hybrid Intelligent Systems. Kluwer Academic Pub., Dordrecht (1995)
7. Corchado, E., Abraham, A., de Carvalho, A.: Editorial: Hybrid intelligent algorithms and applications. Inf. Sci. **180**, 2633–2634 (2010)
8. Herrero, Á., Corchado, E., Pellicer, M.A., Abraham, A.: Movih-ids: A mobile-visualization hybrid intrusion detection system. Neurocomp. **72**, 2775–2784 (2009)
9. Tapia, D.I., Fraile, J.A., de Luis, A., Bajo, J.: Healthcare information fusion using context-aware agents. In: HAIS, pp. 96–103 (2010)
10. Villar, J., de la Cal, E., Sedano, J.: Hybrid Artificial Intelligence Systems: Minimizing energy consumption in heating systems under uncertainty. In: Corchado, E., Abraham, A., Pedrycz, W. (eds.) HAIS 2008. LNCS, vol. 5271, pp. 583–590 Springer, Heidelberg (2008)
11. Cucchiara, R., Fornaciari, M., Prati, A., Santinelli, P.: Mutual calibration of camera motes and rfids for people localization and identification. In: ACM/IEEE ICDSC, pp. 1–8 (2010)
12. Luckham, D.: The Power of Events: An Introduction to Complex Event Processing in Distributed Enterprise Systems. Addison-Wesley Longman, Amsterdam (2002)
13. Paschke, A.: A homogenous reaction rule language for complex event processing. In: EDA-PS (2007)
14. Dubois, D., Prade, H., Smets, P.: New semantics for quantitative possibility theory. In: ECSQARU, pp. 410–421 (2001)
15. Jousselme, A.L., Grenier, D., loi Boss: A new distance between two bodies of evidence. Information Fusion **2**, 91–101, (2001)
16. Smets, P.: Analyzing the combination of conflicting belief functions. Inf. Fusion **8**, 387–412 (2007)