Fusion of Audio and Temporal Multimodal Data by Spreading Activation for Dweller Localisation in a Smart Home

Pedro Chahuara, François Portet, Michel Vacher

Fusion of Audio and Temporal Multimodal Data by Spreading Activation for Dweller Localisation in a Smart Home*

Pedro Chahuara, François Portet, Michel Vacher
LIG UMR 5217, UJF-Grenoble 1 / Grenoble INP / UPMF-Grenoble 2 / CNRS,
Laboratoire d'Informatique de Grenoble
Grenoble, F-38041, FRANCE
{pedro.chahuara,francois.portet,michel.vacher}@imag.fr

Abstract

In this paper, an approach to locate a person using non visual sensors in a smart home is presented. The information extracted from these sensors gives uncertain evidence about the location of a person. To improve robustness of location, audio information (used for voice command) is fused with classical domotic sensor data using a two-level dynamic network and using an adapted spreading activation method that considers the temporal dimension to deal with evidence that expire. The automatic location was tested within two different smart homes using data from experiments involving 25 participants. The preliminary results show that an accuracy of 90% can be reached using several uncertain sources.

1 Introduction

The objective of this work is the continuous location of an inhabitant in their home using sources that are non-visual (i.e., without camera) and indirect (the person does not wear a sensor). It is part of the Sweet-Home (`http://sweet-home.imag.fr/`) project which aims to design an intelligent controller for home automation through a voice interface for improved comfort and security. In this vision, users can utter vocal orders from anywhere in their house, thanks to microphones set into the ceiling. It is thus particularly suited to assist people with disabilities and the growing number of elderly people in living autonomously as long as possible in their own home. Within the smart home domain, this concept is known as *Aging-In-Place* [Marek and Rantz, 2000] and consists in allowing seniors to keep control of their environment and activities, to increase their autonomy, well-being and their feeling of dignity.

Among the main data processing tasks a smart home must implement, detecting the correct location of the person plays a crucial role to make appropriate decisions in many applications (e.g., home automation orders, heating and light control, dialogue systems, robot assistants) and particularly for health and security oriented ones (e.g., distress call, fall, activity monitoring). For instance, in the Sweet-Home context, if the person says "turn on the light", the location of the person and the lamp which is referred to must be deduced. However, in the context of a vocal command application, noise, reverberation and distant speech can alter the recognition quality and lead to incorrect inferences [Vacher *et al.*, 2011]. To improve the robustness of the automatic location, we propose to combine audio source with other sources of information.

Automatic location becomes particularly challenging when privacy issues prevent the systematic use of video cameras and worn sensors. In the Sweet-Home project, only classical home automation and audio sensors are taken into account. These sensors —Presence Infra-red Detector (PID), door contacts, and microphones— only inform indirectly and transiently about the location of a person. Automatic location is thus a challenging task that must deal with indirect, uncertain and transient multiple sources of information.

In this paper, we present a new method developed for automatic dweller location from non-visual sensors. After a brief state of the art of location techniques in Section 2, the approach we adopted to locate a person is presented in Section 3. It is based on a fusion of information obtained from various sensors (events) through a dynamic network that takes into account the previous activations and the uncertainty of the events. The adaptation of the method to two smart homes is described in Section 4 and the results of the experiments are summarised in Section 5. The paper ends with a brief discussion of the results and gives a future work outlook.

2 Location of an Inhabitant: Common Techniques in Smart Homes

Techniques for locating people in a pervasive environment can be divided into two categories: those that use sensors explicitly dedicated to this task and worn by people such as a GPS bracelet, and those that use sensors that only inform implicitly about the presence of a person in a confined space such as infrared presence sensors or video cameras.

The wearable sensors are often used in situations where the person has a social activity (museum visit) or for professional, health or safety reasons (e.g., patients with Alzheimer's disease running away). Despite their very good location performance, they are not adapted to an informal and comfortable

*This work is part of the Sweet-Home project founded by the French National Research Agency (Agence Nationale de la Recherche / ANR-09—VERS-011)

in-home use. Indeed, these sensors can be awkward and annoying and, except for passive sensors (e.g., RFID), they require the systematic checking of the batteries. Moreover, if the goal is to improve the daily living comfort, the constraint of a wearable sensor may be a strong intrusion into the intimate life. That is why this paper focuses on techniques using environmental sensors (video, sound, motion sensor, door contacts, etc..).

Video analysis is a very interesting modality for home automation which is used in many projects [Marek and Rantz, 2000; Moncrieff *et al.*, 2007]. However, video processing requires high computational resources and can be unreliable and lacking in robustness. Moreover, installing video cameras in a home may be perceived as too much intrusion into intimate life, depending on the kind of video processing that is installed (e.g., plain vs. silhouette based video processing or hiding [Moncrieff *et al.*, 2007]).

Another usual source of localization can be derived from household appliances and surveillance equipment. For instance, infrared sensors designed for automatic lighting were used to evaluate the position and the activity of the person [Le Bellego *et al.*, 2006; Wren and Tapia, 2006]. The use of some devices can also be detected using new techniques that identify the signatures of an electrical appliance on the household electric power supply [Berenguer *et al.*, 2008].

Another interesting modality in home automation is the analysis of the audio channel, which, in addition to providing a voice command, can bring various audio information such as broken glass, slamming doors, etc. [Vacher *et al.*, 2010]. By its omnidirectional or directional nature, the microphone is a promising sensor for locating events with a high sensitivity or high specificity. There is an emerging trend to use such modality in pervasive environment [Bian *et al.*, 2005; Moncrieff *et al.*, 2007; Vacher *et al.*, 2010]. Audio sources require far less bandwidth than video information and can easily be used to detect some activities (e.g., conversations, telephone ringing). However, if the video is sensitive to changes in brightness, the audio channel is sensitive to environmental noise. The audio channel, while a relevant and affordable modality is therefore a noisy source and sometimes highly ambiguous.

Throughout this state of the art, it appears that no source taken alone makes a robust and resource-efficient location possible. It is therefore important to establish a location method that would benefit from the redundancies and complementarities of the selected sources. There is a large literature in the domain of activity recognition on such methods mainly using probabilistic graph-based methods such as Bayesian networks [Dalal *et al.*, 2005; Wren and Tapia, 2006] or Markov models [Wren and Tapia, 2006; Chua *et al.*, 2009]. However, given the large number of sensors, building HMM models taking all the transition states into account for really time processing would be extremely costly. More flexible graph-based approaches based on sensor network that include a hierarchy of processing levels (Bayesian and HMM classifiers) were proposed [Wren and Tapia, 2006]. However, if temporal order is often taken into account, the temporal information about duration or absolute date is rarely considered in these models .

Recently, Niessen et al. [Niessen *et al.*, 2008] proposed to apply dynamic networks to the recognition of sound events. In their two-level network, the input level is composed of sound events, the first level represents the assumptions related to an event (e.g., ball bounce or hand clap), and the second level is the context of the event (e.g., basketball game, concert, play). Each event activates assumptions according to the input event and the contexts to which these assumptions are linked. These assumptions then activate the contexts, reenforcing them or not. Thus, there is a bidirectional relationship between contexts and assumptions. For instance, if several previously recognized sounds are linked to a concert context, the next sound will be more likely related to a concert context. The method imposes no pattern but the notion of time is explicitly taken into account by a time constant that reduces the importance of an event according with its age. Given the flexibility provided by this approach, we chose to adapt it to the location of a person in a flat using multisource information.

3 Location of an Inhabitant by Dynamic Networks and Spreading Activation

The method developed for locating a person from multiple sources is based on the modelling of the links between observations and location assumptions by a two-level dynamic network. After a brief introduction to dynamic networks and spreading activation, the method adapted to work with multiple temporal sources is described.

Dynamic Networks and Spreading Activation

The spreading activation model, is employed in AI as a processing framework for semantic or associative networks and is particularly popular in the information retrieval community [Crestani, 1997]. Briefly, the considered network $\mathcal{N}$ is a graph where nodes represent concepts and where arcs, usually weighted and directed, represent relationships between concepts. The activation process starts by putting some 'activation weight' at an input node that spreads to the neighbouring nodes according to the strength of their relationships and then spreads to the other neighbours and so on until a termination criteria is satisfied. The activation weight of a node is a function of the weighted sum of the inputs from the directly connected nodes. For detailed introduction to spreading activation, the reader is referred to [Crestani, 1997]. Within this model, dynamic network have also bee proposed to represent knowledge that evolves with time [Niessen *et al.*, 2008]. A network is dynamic in the sense that it changes according to inputs that can modify the structure and/or the strength of the relationships between nodes. The spreading activation in a dynamic network provides a flexible and intuitive framework to represent associations between concepts which is particularly interesting to fuse evidence that decay. However, to the best of our knowledge, few approaches have focused on the case of temporal sources whose activation decreases with time [Niessen *et al.*, 2008]. In the following, we present the temporal multisource approach to locate a person in an flat.

Temporal Dynamic Networks for Multisource Fusion

The dynamic network that we designed is organized in two levels: the first level corresponds to location hypotheses generated from an event; and the second level represents the occupation context for each room whose weight of activation indicates the most likely location given the previous events. Location hypotheses correspond to area where the person can be at a specific time while occupation contexts correspond to rooms in which the person is over time. Our approach uses the following definitions:

Definition 1 (Observation) *An observation o_n is a data structure generated when a sensor reacts to the event e_n at time $t_n \in \mathbf{R}^+$ with $n \in \mathbf{N}$. Each observation is related to a sensor $o.sensor$ and has a sensor type $o.type$.*

Definition 2 (Simultaneous observations) *Two observations o_n^i and o_k^j are simultaneous if $t_k \in [t_n - d, t_n + d]$, with $d \in \mathbf{R}^+$ a predefined delay.*

Definition 3 (Observation activation) *The activation $A_n^o \in [0,1]$ of an observation o_n represents the intensity of the evidence being integrated into the network. It can be derived from the weight or probability of the classifier/detector generating the observation. In our case, A_n^o is based on its ambiguity such that for a set of simultaneous observations of same type O_n, $\sum_{o \in O_n} A_n^o = 1$.*

Definition 4 (Location hypothesis) *$h_n^i \in L$, where $L = \{Loc_1, \ldots, Loc_R\}$ is the hypothesis that the inhabitant is at location i at time t_n. These hypotheses are created only from the observations at time t_n.*

Definition 5 (Occupation context) *$c^i \in R$ where $R = \{Room_1, \ldots, Room_S\}$ is the occupation context of the i^{th} room.*

Definition 6 (Relationship weight) *$w \in [0,1]$ is the importance of the relationship between two nodes in the network. w_{o,h^i} is the weight between an observation and the i^{th} hypothesis whereas w_{h^i,c^j} is the weight between the i^{th} hypothesis and the j^{th} context.*

Definition 7 (Decay function) *The decay function $f(t_n, t_{n-1}) = e^{-\frac{\Delta_t}{\tau}}$, with $\Delta_t = t_n - t_{n-1}$ represents the decrease of the context through time. It makes it possible to keep a short-term memory about contexts.*

The dynamic network evolution is summarised by the following algorithm:

1. for every new observation o_n^k, a new node is created;
2. thereupon hypothesis nodes h_n^i are created and connected to o_n^k with weights w_{o^k,h^i};
3. hypothesis nodes h_n^i are connected to occupation context nodes c^j with weights w_{h^i,c^j};
4. activation spreads from o_n^k to h_n^i and the activation of each h_n^i is calculated;
5. activation spreads from h_n^i to c^j and the activation of each c^j is recalculated;
6. the node c^j with the highest activation becomes the present location;

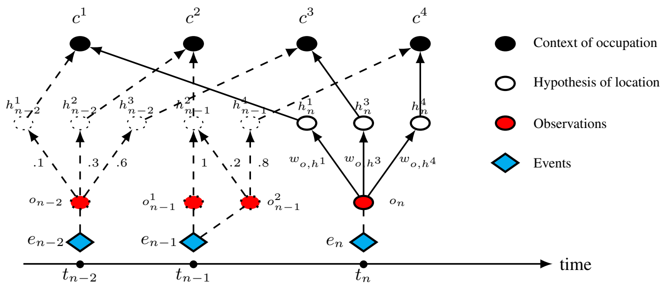


Figure 1: Example of Dynamic Network

7. all the nodes h_n^i and o_n^k are deleted from the network.

An example of dynamic network is shown in Figure 1. At time t_{n-2}, the event e_{n-2} is detected by a sensor which generates the observation o_{n-2} from which 3 hypotheses are derived : h_{n-2}^1 with a relationship weight of 0.1 towards the context c^1, h_{n-2}^2, with weight 0.3 towards c^2 and h_{n-2}^3 with 0.6 towards c^3. If no previous events occurred, then c^3 would be the most probable location. At time t_{n-1}, two simultaneous observations caused by event e_{n-1} are integrated within the network. Every node created previously (i.e., at t_{n-2}) is discarded, except the contexts which are always kept. Active contexts are weighted by $f(t_{n-1}, t_{n-2})$ and the activation of the hypothesis h_{n-1}^2 is added to c^2 and h_{n-1}^4 is added to c^4. The method is applied subsequently at time t_n.

Spreading Activation

The activation of a node is typically defined [Crestani, 1997] by the formula $n_i(t) = \sum_{i \neq j} w_{i,j} \times A^j(t)$ where $w_{i,j}$ is the weight, j corresponds to a neighbour of i and $A^j(t)$ is the activation of its neighbours at time t. A node that has been activated by a neighbour node cannot spread its activation back to it. In our case, activations are always triggered by an observation o with a bottom-up spreading. Once the accumulated activation from neighbours $n(t)$ is obtained, the output activation of the node is calculated. It differs according to the node level. For location hypotheses, the activation $A_n^{h^i} \in [0,1]$ is computed using Formula 1:

$$A_n^{h^i} = n_i(t_n) = \sum_{o \in O_{t_n}} w_{o,h^i} \, A_n^o \tag{1}$$

Regarding the occupation contexts, the output activation results from the previous activation weighted by the decay function and the accumulated activation of hypotheses. Equation 2 describes the activation of an occupation context A^{c^i} as a consequence of an external activation at time t_n.

$$A^{c^i}(t_n) = M \times A_n^{h^i} + e^{-\frac{\Delta t}{\tau}} A^{c^i}(t_n - \Delta t) \times [1 - A_n^{h^i}] \tag{2}$$

where $A^{c^i}(t_n - \Delta t)$ is the previous activation, $M = 1$ is the maximal activation and $e^{-\frac{\Delta t}{\tau}}$ is the decay function. Therefore, if no event appears during $5 \cdot \tau$ seconds, the contexts activation can be considered zero. The introduction of M constrains the activation value between 0 and 1.

Computation of the node level Relationship
Given that the network is composed of two layers, two types of relationship exist: *Observation-Hypothesis* and *Hypothesis-Context*. The links between the different layers depend strongly on the application and the environment considered.

The **Hypothesis-Context** relationship is in our case of type one-to-one because a hypothesis of location is only related to a unique room. It is an experimental choice since some hypotheses about rooms loosely separated could activate several occupation contexts. Thus $w_{h^i,c^j} = 1\,\forall i = j,\ 0$ otherwise.

The **Observation-Hypothesis** relationship is unidirectional and of type one-to-many. Weights and hypotheses vary depending on the observations and prior knowledge about this relationship. In order to include this prior knowledge in the network, the relationship weight is defined by formula 3 in the form of probabilities where the relationship weight between the current (possibly set of simultaneous) observation(s) O_n and hypothesis h^i is defined by the probability of observing the inhabitant at location i given the current observation(s) and the context $\mathcal{C}$.

$$w_{o,h^i}(t_n) = P(loc = i \mid O_n, \mathcal{C}) \quad (3)$$

4 Adaptation of the Method to Pervasive Environments

Two pervasive environments were considered in our study: the DOMUS smart home and the Health Smart Home(HIS) of the Faculty of Medicine of Grenoble. Every experiment in these smart homes considered only one inhabitant at a time. In section 4.1 details of both corpora are given, then sections 4.2 and 4.3 explain how relationships between layers are computed for each smart home and which *a priori* information is taken into account to derive the inhabitant's location.

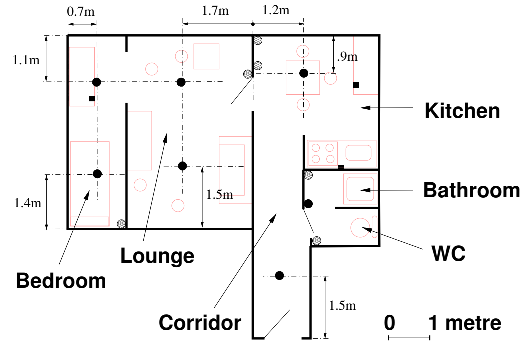
4.1 Pervasive Environments and Data Used

The HIS corpus was acquired during experiments [Fleury *et al.*, 2010] aiming at assessing the automatic recognition of Activities of Daily Living (ADL) of a person at home in order to automatically detect loss of autonomy. Figure 2a describes the 6-room Health Smart Home of the Faculty of Medicine of Grenoble at the TIMC-IMAG laboratory [Le Bellego *et al.*, 2006]. The data considered in this study consisted of about 14 hours of 15 people recordings using the following sensors:

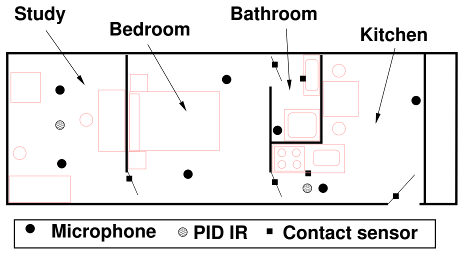
- 7 microphones (Mic) set in the ceiling;
- 3 contact sensors on the furniture doors (DC) (cupboards in the kitchen, fridge and dresser in the bedroom);
- 6 Presence Infrared Detectors (PID) set on the walls at about 2 metres in height.

The Sweet-Home corpus was acquired in realistic conditions, using the DOMUS smart home. This smart home was designed and set up by the Multicom team of the Laboratory of Informatics of Grenoble. Figure 2b shows the details of the flat. The data considered in this study consisted of about 12 hours of 10 people recordings performing daily activities using the following sensors:

- 7 microphones (Mic) set in the ceiling;



(a) Health smart home



(b) DOMUS Smart Home

Figure 2: Layout of the smart homes used and position of the sensors.

- 3 contact sensors on the furniture doors (DC) (cupboards in the kitchen, fridge and bathroom cabinet);
- 4 contact sensors on the 4 indoor doors (IDC);
- 4 contact sensors on the 6 windows (open/close);
- 2 Presence Infrared Detectors (PID) set on the ceiling.

4.2 Weight computation for the HIS

$w_{o,h^i}(t_n)$, was computed differently for each kind of sensors. For observation o with $o.type \in \{DC, PID\}$ a single hypothesis node is created with a weight $w_{o,h} = 1$. Indeed, spatial informations about PID an DC are unambiguous and certain. For example, opening the fridge can only occurs if the inhabitant is in the fridge area. For both PID an DC, the activation of observations is $A_n^o = 1$.

The microphones cannot be treated in the same manner. Microphones can detect theoretically all the acoustic waves produced in the home making this information highly ambiguous. However, it is possible to estimate from the position of the Mic the areas they can most likely sense and thus the location they are most related to. To take this into account, formula 3 was approximated with $w_{o,h^i}(t_n) = P(loc = i|Mic = j)$ that is the probability that the inhabitant is in the i^{th} room given an observation o_n generated by the j^{th} microphone. To acquire this *a priori* knowledge, two approaches were tested: a naïve approach and a statistical one. Succinctly, for the naive approach, the circle outside which the loss of energy is greater than 75% is considered for each Mic. The weight is calculated as the surface of the intersection between the circle and the rooms with a penalty of 2 when the circle goes beyond a wall. The statistical approach acquired

$P(Loc|Mic)$ estimated by the naïve approach

Mic	1	2	3	4	5	6	7
bedroom	.14	.07	.70	.85			
lounge	.86	.93	.27	.14	.01	.13	.03
kitchen			.03	.02	.10	.87	.50
bathroom					.06		.18
wc					.06		.18
corridor					.77		.10

$P(Loc|Mic)$ estimated from corpus

Mic	1	2	3	4	5	6	7
bedroom	.28	.29	.42	.43	.25	.18	.20
lounge	.59	.56	.47	.41	.07	.07	.09
kitchen	.05	.08	.06	.09	.45	.63	.37
bathroom	.06	.05	.04	.04	.09	.04	.10
wc	.01	.01	.01	.01	.12	.05	.21
corridor		.02	.01	.02	.03	.03	.02

Table 1: Estimation of $P(Loc|Mic)$ for HIS smart home

the probabilities from the annotated corpus. Table 1 shows the weights obtained for both approaches. Apart from this static information, dynamic information, such as the signal energy, was taken into account in case of simultaneous observations on the Mic. The Mic activation, it was computed using the signal-to-noise ratio (SNR) estimated in real-time. Given that A_n^o summarises the observation ambiguity, it was computed as $A_n^o = o.snr / \sum_{obs \in O_n} obs.snr$ where O_n is the set of simultaneous observation at time n.

The fusion of prior and dynamic information provides a better disambiguation. For example, for the HIS flat if two simultaneous observations are detected by the microphones in the kitchen and bathroom with a similar SNR of $12dB$, the formula 1, and the prior information from the naive estimation give as activation $A^{h^{kitchen}} = .87 \times A^{o^6} + .5 \times A^{o^7} = .69$ which is higher than the bathroom activation $A^{h^{bath}} = .09$ even when the SNR is similar.

4.3 Weight computation for Sweet-Home

$w_{o,h^i}(t_n)$ was computed in the same way as for the HIS for DC and PID. However, conversely to contact sensors on furniture and windows which are always linked to a unique hypothesis, contact sensors on the indoor doors (IDC) can be ambiguous regarding the location. The problem is to decide which of the two rooms around the door should have the highest weight. In that case, formula 3 was approximated with the conditional probability $w_{o,h^i} = P(loc = i \mid o.sensor, o.state, \mathcal{C})$, where $o.state \in \{Open/Close\}$ and $\mathcal{C}$ is the inhabitant's location at time t_{n-1}. This *a priori* knowledge was statistically acquired from an annotated corpus different from the one used in the test.

Results of the conditional probabilities estimation indicate that in 97% of cases when a door is open from a room then a transition to the contiguous room is produced, whereas when the door is closed the transition is less certain (66% of cases).

5 Experimentation

For each participant's record, the events from DC, PID and Mic were used to activate a dynamic network to estimate the location of the inhabitant. Location performance was evaluated every second by comparing the context of the highest weight to the ground truth. If they matched, then it was a true positive (TP), otherwise it was a confusion. The accuracy was given by $Acc = nb(TP)/nb(test)$ where $nb(test)$ corresponds to the duration of the record in seconds and $nb(TP)$ the number of seconds in which a TP was obtained.

Sensor and prior information	PID	DC	Mic+ DC	PID+ DC	PID+ Mic	PID+ Mic+DC
SH no prior info.	62.9	59.9	63.7	71.6	64.5	**73.2**
SH prior info.	62.8	73.3	77.4	81.7	64.6	**84.0**
HIS no prior info.	88.9	26.5	32.8	**89.4**	87.7	88.2
HIS prior naive info.	88.9	26.5	34.1	89.4	89.0	**89.5**
HIS prior stat. info.	88.9	26.5	34.8	89.4	89.7	**90.1**

Table 2: Accuracy with several combinations of sources

For Sweet-Home a first experiment was done without using the prior probabilistic knowledge about room doors contact sensors, afterwards the method was executed using these probabilities to evaluate how significant their contribution is. Likewise, three independent experiments were carried out with the HIS corpus: without a priori ($P(loc|Mic) = 1$ when the microphone was in the room, 0 otherwise), with the naïve approach and with the statistical approach. Table 2 shows the results of both corpora using several combinations of sensor.

In the case of Sweet-Home, it is clear that the fusion of information improved the accuracy since it rises as more sensors information is combined. Even when the precision of infrared sensors was good, the overall results of the method using only these sensors was low (63%) as only two of them were set in the 4-room flat. This led to a poor sensitivity. In the second row, the accuracy using the information of door contact on room doors is reported. It can be noticed that the learned probabilities had a significant impact on the performance. Here, the column DC includes also the results with IDC. In every case, *a priori* information about IDC had a positive impact on the performances, confirming that prior knowledge can be used to improve the performance.

From HIS experiement, it can be noticed that, in some cases, fusion of information did not improve the accuracy. The door contact information slightly improved the accuracy compared to that obtained only with the infrared sensors. On the other hand, adding the sound information decreased the performance (88.2 % versus 89.4 %). One reason for this may the high level of confusion between sound and speech of the AUDITHIS system which reached 25% of classification errors. Nevertheless, once again an increase of performance was achieved by means of the prior knowledge introduction: results are better or similar in every combination of sensors when using the probabilities. There is a slight advantage of statistical approach over the naïve one but the naïve approach does not require any dataset to be acquired and thus simplifies the set up in new pervasive environements.

6 Discussion and Perspectives

The results showed that the information fusion by spreading activation is of interest even when the sources have very good accuracy. It is the case for infrared sensors (but with imperfect sensitivity) and for door contact sensors. The use of less certain localisation sources, such as speech recognition, can then improve performance in many cases. Another important finding is that *a priori* knowledge about sensors is a possible leverage to gain a higher accuracy as it was done with the contact sensors of the room doors for the Sweet-Home corpus and the microphones for the HIS corpus. In those cases, the introduced knowledge was expressed in terms of conditional probabilities and its exploitation was demonstrated to be useful. Furthermore, the approach is general enough to include different kinds of ambiguous sensor as input to the dynamic network. Given the source, the probability of the inhabitant being in a room given some feature of the sensor data can be estimated and this prior knowledge can be applied to enhance localisation. This is the case, for instance, of the water meter. Even if this information cannot be directly used for localisation, it is feasible to estimate the probability about the inhabitant's location given the change of flow rate in order to use this probability when generating hypotheses in the dynamic network.

Several ways to improve this method can be followed. So far, besides direct information given by sensors, we have applied some knowledge based on specific sensors features. However, it would be advantageous to use other characteristics of the environment. One way could be to use the topology of the flat as in [Wren and Tapia, 2006]. For instance, an occupant can not move from the bedroom to the front door without going through the lounge, etc. Further extensions of our method include Markovian techniques to estimating the probability of the present inhabitant's location given their precedent location. We believe that it could be a relevant contribution when fusioned with the sources of information already described in this work. The next step is to apply this method to classify the sounds of everyday life using the location context to disambiguate the sound classification, and to test the general suitability of the approach by confronting the system to actual users (elderly and frail people).

References

[Berenguer *et al.*, 2008] M. Berenguer, M. Giordani, F. Giraud-By, and N. Noury. Automatic detection of activities of daily living from detecting and classifying electrical events on the residential power line. In *HealthCom'08, 10th IEEE Int. Conf. on e-Health Networking, Applications & Service*, 2008.

[Bian *et al.*, 2005] Xuehai Bian, Gregory D. Abowd, and James M. Rehg. Using sound source localization in a home environment. In *Third International Conference of Pervasive Computing*, pages 19–36, 2005.

[Chua *et al.*, 2009] Sook-Ling Chua, Stephen Marsland, and Hans W. Guesgen. Spatio-temporal and context reasoning in smart homes. In *Proceedings of the COSIT Workshop on Spatial and Temporal Reasoning for Ambient Intelligence Systems*, 2009.

[Crestani, 1997] F Crestani. Application of spreading activation techniques in information retrieval. *Artificial Intelligence Review*, 11(6):453–482, 1997.

[Dalal *et al.*, 2005] S Dalal, M Alwan, R Seifrafi, S Kell, and D Brown. A rule-based approach to the analysis of elders activity data: Detection of health and possible emergency conditions. In *AAAI Fall 2005 Symposium*, 2005.

[Fleury *et al.*, 2010] Anthony Fleury, Michel Vacher, François Portet, Pedro Chahuara, and Norbert Noury. A multimodal corpus recorded in a health smart home. In *LREC Workshop Multimodal Corpora and Evaluation*, pages 99–105, Matla, 2010.

[Le Bellego *et al.*, 2006] G. Le Bellego, N. Noury, G. Virone, M. Mousseau, and J. Demongeot. A model for the measurement of patient activity in a hospital suite. *IEEE Transactions on Information Technologies in Biomedicine*, 10(1):92 – 99, 2006.

[Marek and Rantz, 2000] KD Marek and MJ Rantz. Aging in place: a new model for long-term care. *Nursing Administration Quarterly*, 24(3):1–11, 2000.

[Moncrieff *et al.*, 2007] Simon Moncrieff, Svetha Venkatesh, and Geoff A. W. West. Dynamic privacy in a smart house environment. In *IEEE Multimedia and Expo*, pages 2034–2037, 2007.

[Niessen *et al.*, 2008] Maria E. Niessen, Leendert van Maanen, and Tjeerd C. Andringa. Disambiguating sounds through context. In *Proceedings of the 2008 IEEE International Conference on Semantic Computing*, pages 88–95. IEEE Computer Society, 2008.

[Vacher *et al.*, 2010] Michel Vacher, Anthony Fleury, François Portet, Jean-François Serignat, and Norbert Noury. *Complete Sound and Speech Recognition System for Health Smart Homes: Application to the Recognition of Activities of Daily Living*, pages 645 – 673. Intech Book, 2010.

[Vacher *et al.*, 2011] Michel Vacher, François Portet, Anthony Fleury, and Norbert Noury. Development of audio sensing technology for ambient assisted living: Applications and challenges. *International Journal of E-Health and Medical Communications*, 2(1):35–54, 2011.

[Wren and Tapia, 2006] Christopher R. Wren and Emmanuel Munguia Tapia. Toward scalable activity recognition for sensor networks. In *Location- and context-awareness*, 2006.