Location of an Inhabitant for Domotic Assistance Through Fusion of Audio and Non-Visual Data

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

Abstract—**In this paper, a new method to locate a person using multimodal non-visual sensors and microphones in a pervasive environment is presented. The information extracted from sensors is combined using a two-level dynamic network to obtain the location hypotheses. This method was tested within two smart homes using data from experiments involving about 25 participants. The preliminary results show that an accuracy of 90% can be reached using several uncertain sources. The use of implicit localisation sources, such as speech recognition, mainly used in this project for voice command, can improve performances in many cases.**

I. Introduction

Several technologies have been used to set up smart environments able to ease the person's life and to provide adequate assistance (e.g., infra-red sensors, video cameras, RFID . . .). Audio processing technology has a great potential to become one of the major human-machine interaction modalities in smart home. It is physically intangible and does not force the user to be at a particular place in order to operate. Moreover, it can provide interaction using natural language so that the user does not have to learn complex computing procedures or jargon. Voice interfaces can be much more suited to disabled people and seniors who have difficulties in moving or seeing, than tactile interfaces (e.g., remote control) which require physical and visual interaction. Moreover, audio processing is particularly adapted to distress situations. A person, who cannot move after a fall but being conscious, can still call for assistance while a remote control may be unreachable. Despite this, few smart home projects have seriously considered audio technology in their design [3], [6], [7]. Part of this can be attributed to the fact that this technology, though mature, still needs to address important challenges [10].

To improve autonomy, comfort and security at home, we are developing a new smart home system called Sweet-Home whose main human-machine interaction modality is based on audio processing technology. Among the first data processing tasks a smart home must implement, automatic locating of the person is essential. Indeed, the location of the person plays a crucial role to make appropriate decisions for 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). Automatic locating becomes particularly challenging when privacy issues prevent the systematic use of video cameras and worn sensors. Another source of localization can be derived from household appliances and surveillance equipment [5], [11]. The analysis of the audio channel is another interesting modality in home automation, which, in addition to providing a voice command, can bring various audio information such as broken glass, slamming doors, etc. [9]. There is an emerging trend to use such modality in pervasive environments [1], [7], [9]. The audio information requires far less bandwidth than video information and can easily 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 [10]. It appears that no source taken alone makes possible a robust and cheap location. It is therefore important to establish a location method that would benefit from the redundancies and complementarities of the selected sources.

In this paper, we present a new method developed for automatic dweller location from non-visual sensors. After an introduction to the Sweet-Home project in Section II, the approach we adopted to locate a person is presented in Section III. This approach was tested within real smart homes and the results of the experiments are described in Section IV. The paper ends with a brief discussion of the results.

II. Sweet-Home: an Audio-Based Smart Home System

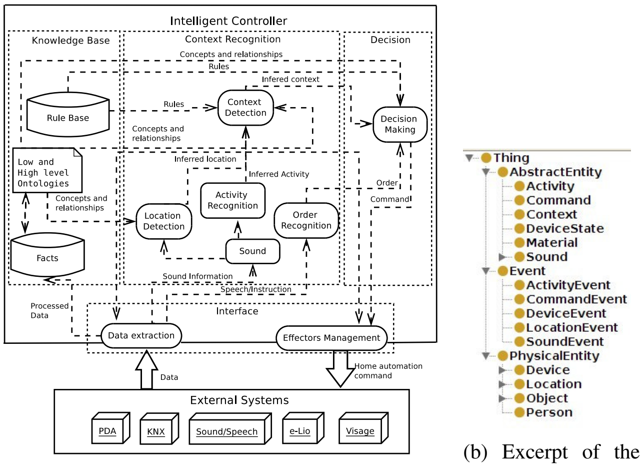
The Sweet-Home project (http://sweet-home.imag.fr) is a French national supported research project aiming at designing a new smart home system based on audio technology focusing on three main aspects: to provide assistance via *natural human-machine interaction* (voice and tactile command), to ease *social inclusion* and to provide *security reassurance* by detecting situations of distress. If these aims are achieved, then the person will be able to pilot their environment at any time in the most natural way possible. To assess the acceptance and fear of this new technology, a qualitative user evaluation was performed. 8 healthy persons between 71 and 88 years old, 7 relatives (child, grand-child or friend) and 3 professional carers have been recruited. During about 45 minutes, they were questioned in co-discovery in a fully equipped smart home alternating between interview and wizard of Oz periods followed by a debriefing. The four important aspects of the project have been assessed: voice command, communication with the outside world, domotic system interrupting a person's activity, and electronic agenda. Succinctly, in each case the voice based solution was far better accepted than more intrusive solutions. Thus, audio technology appears to have a great

PervasiveHealth 2011, May 23-26, Dublin, Republic of Ireland

DOI 10.4108/icst.pervasivehealth.2011.246054

potential to ease daily living for elderly and frail persons.

The input of the Sweet-Home system is composed of the information from the domotic system transmitted via a local network and information from the microphones transmitted through radio frequency channels. The Sweet-Home system will be piloted by an intelligent controller which will capture all streams of data, interpret them and execute the required actions. The diagram of this intelligent controller is depicted in Figure 1a. The knowledge of the controller is defined using two semantic layers: the *low-level* and the *high-level* ontologies. The former ontology is devoted to the representation of raw data and network information description. State, location, value and URI of switches and actuators are examples of element to be managed by the I/O processors. The high level ontology, whose taxonomy is shown in Figure 1b, represents concepts being used at the reasoning level. These concepts are organized in 3 main branches: the Abstract Entity represents the different actions that can be performed in a home and the context in which a home can be (e.g., making coffee, being late), the Physical Entity represents things that are present in the home (e.g., the dweller, electrical devices), and finally, the Event concept represent the transient observations of one abstract entity involving zero or several physical entity (e.g., at 12:03 the dweller is sleeping). This separation between low and high levels makes possible a higher re-usability of the reasoning layer when the sensor network and the home have to be adapted [4]. The estimation of the current context is carried out through the collaboration of several processors, each one being specialized in a certain context aspect, such as location detection or activity recognition. All processors share the knowledge specified in both ontologies and use the same repository of facts. Furthermore, the access to the knowledge base is executed under a service oriented approach that allows any processor being registered to be notified only about particular events and to make inferred information available to other processors. This data and knowledge centred approach ensures that all the processors are using the same data structure and that the meaning of each piece of information is clearly defined among all of them. Once the current context has been determined, the controller evaluates if an action must be taken, such as making an emergency call in case of a circumstance of distress.



(a) The Intelligent Controller Diagram

(b) Excerpt of the high level ontology

Fig. 1: Intelligent Controller diagramm and excerpt of the Sweet-Home ontology.

III. 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 events and location assumptions by a two-level dynamic network. Recently, Niessen et al. [8] presented an approach based on a two-level dynamic networks to disambiguate the recognition of sound events. The input level is composed of sound events, level one represents the assumptions related to an event (e.g., ball bounce or hand clap), and level two 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.

We adapted this method to multisource fusion with the aim of locating a person. The dynamic network that we designed is organized into two levels: the first level corresponds to location hypotheses generated from an event; and the second level represents the contexts for which the activation indicates the most likely location given the previous events. Thus, the activation A_i^h of the i^{th} hypothesis depends exclusively on the probability given by the event e at time t_n which generated it. It is computed using formula 1.

$$A_i^h(t_n) = P(location = i \mid e_{t_n}) \qquad (1)$$

And the activation A_i^c of the i^{th} context depends on its previous value and its associated hypothesis. It is computed using formula 2.

$$A_i^c(t_n) = e^{-\frac{\Delta t}{\tau}} A_i^c(t_n - \Delta t) + A_i^h(t_n) \qquad (2)$$

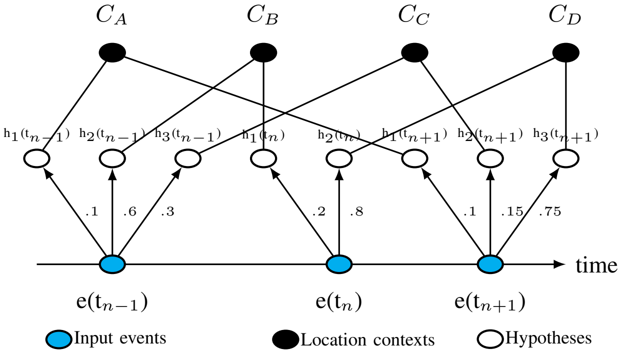


Fig. 2: Example of a Dynamic network

Figure 2 gives an example of activation for this network. At time t_{n-1}, the event $e(t_{n-1})$ appears and generates 3 hypotheses: $h_1(t_{n-1})$ with a weight of 0.1 towards context

C_A, $h_2(t_{n-1})$ with a weight of 0.6 towards context C_B and $h_3(t_{n-1})$ with a weight of 0.3 towards context C_C. If there was no prior event, context C_B will be the most certain. At time t_n, the previous weights will be weighted by $e^{-\frac{t_n - t_{n-1}}{\tau}}$ to which will be added the weights generated by the assumptions related to the event $e(t_n)$: $h_1(t_n)$ and $h_2(t_n)$ towards contexts C_B and C_D respectively. The introduction of the time constant permits us to estimate the certainty of finding a person in a room according to its latest location. The decrease of this certainty is implemented by the forgetting function $e^{-\frac{\Delta t}{\tau}}$. The method will be applied in the same way at time t_{n+1} when the context C_D will receive the greatest activation and will be selected.

IV. Experimentation

A. Pervasive Environments and Data Used

The approach was tested on corpra acquired in two fully equipped smart homes: the Sweet-Home corpus and the HIS corpus. The HIS corpus was acquired during experiments [2] 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. 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); and 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 described in Figure 3. This smart home was designed and set up by the Multicom team of the Laboratory of Informatics of Grenoble to observe users' activities interacting with the ambient intelligence of the environment. The data considered in this study consisted of about 12 hours of 10 people recordings performing daily activities using the following sensors: 7 Mics; 3 DCs on the furniture doors; 4 contact sensors on the indoor doors (IDC); 4 DCs on the windows; and 2 PIDs set on the ceiling.

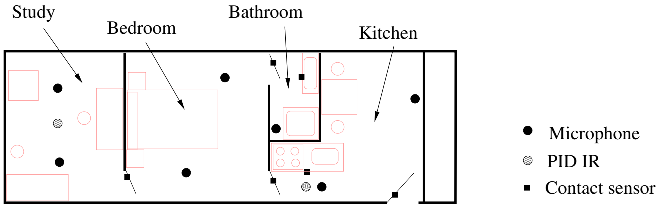


Fig. 3: Layout of the DOMUS smart home and position of the sensors.

B. Weight Estimation

The *hypothesis-context* relationship is, in our case a one-to-one relation because a location hypothesis is related to only one room. It is a experimental choice as some assumptions on weakly separated rooms (e.g., lounge/bedroom) could activate several contexts. The weight of this relationship is always 1.

The *event-hypothesis* relationship is unidirectional and one-to-many. The weights and the hypotheses generated vary according to the source of the event and to the topology of the home. In the case of contact sensors on the furniture and on the windows as well as in the case of the PID, the event was linked to only one hypothesis. Indeed, the DCs and PID delivered non-ambiguous information, thus an event activated with a weight of 1 the hypothesis of the room in which it was situated. For the case of the IDCs of the Sweet-Home corpus, two use cases were considered : the door opening, and the door closing based on the hypothesis that when a door is being open it is more probable that it implies a change of room while when a door is closed the change of room is not certain. Thus, conditional probabilities were computed using the data of 5 participants which were not used in the final test data. When a new DC event arrives at time t_n, the weight W for room $Room$ is given by $W(Room) = P(Room \mid DC, S, Context)$, where $DC, Room \in \{Study, Bedroom, Kitchen, Bathroom\}$, the state $S \in \{Open/Close\}$ and the context $Context = \arg\max_{Room} P(Room)$ at t_{n-1}. These values were calculated for all the combinations of the variables DC,S and $Context$. Then, during the execution of the method, weight estimation was performed dynamically using the learned data. Results of the conditional probabilities estimation indicate that most of the time (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). It seems to support the hypothesis of the two use cases mentioned above. For sound events, the hypothesis weights were computed dynamically. A sound event is generally part of a set of events $E = \{e_j\}$ detected simultaneously on multiple microphones. In this data set, only the sounds classified as speech were considered. Hypotheses were generated from the rooms where the sound events occurred. The weights of the relations event-hypothesis were computed for each location hypothesis p_i using the Signal-to-Noise Ratio (SNR) of each detection with $P(p_i \mid E) = \sum_{e_j \in E \wedge e_j.room = p_i} snr_l(e_j) / \sum_{e_j \in E} snr_l(e_j)$ where e_j is an atomic sound event, p_i is the i^e room and $e_j.room$ (resp $e_j.snr$) is the room where the event was observed (resp. SNR). Thus, the sound event with the highest SNR generates the most likely hypothesis.

C. Results

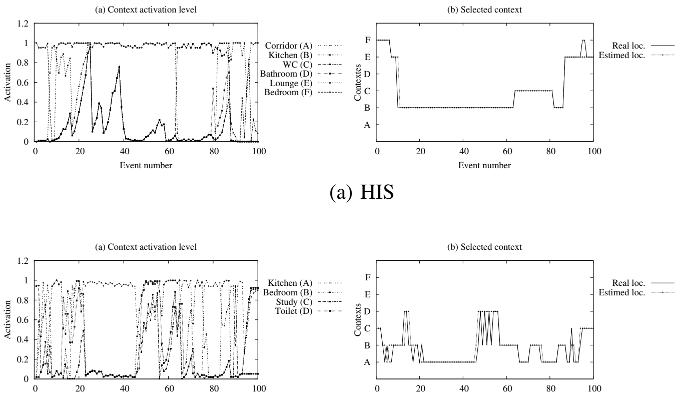
Sensor	PID	DC	Mic+ DC	PID+ DC	PID+ Mic	PID+ Mic+DC
glob. acc. Sweet-Home (%)	63	73	77	82	65	**84**
S-H no DC on doors (%)	63	60	64	72	65	**73**
glob. acc. HIS (%)	95	39	44	**96**	91	92
participant #10 HIS	60	31	78	61	**97**	**97**

TABLE I: Accuracy with several combinations of sources

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 results were summarised in a confusion matrix from which the accuracy Acc was computed by $Acc = \frac{nb(TP)}{nb(test)}$ where $nb(test)$ corresponds to the duration of the record in seconds. Table I shows the results of both corpus.

For Sweet-Home it is clear that the fusion of information improved the accuracy since it rises as more sensors information are combined. Even when the precision of infrared sensors was good, the overall results of the method using only these sensors is low (63%) as only two of them have been set in the 4-room flat. This led to a poor sensitivity. In the second row, the accuracy without using the information of door contact on room doors is reported. It can be noticed that the learned probabilities have a significant impact on the performance. In every case, DC on room doors had a positive impact on the performances. From the results on HIS corpus, 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 (91 % versus 96 %). One reason for this may the high level of confusion between sound and speech of the AUDITHIS system [9] which reached 25% of classification errors.

Nevertheless, the sound information was useful to improve localisation in some cases. Indeed, there were situations where the room change was not detected by the PID while speech was well identified, which compensated for the low infrared sensitivity. The last line of Table I shows the results for participant #10. In this case, it is clear that each source taken alone did not lead to good location but that the combination of sources provided a clear performance gain (60 % to 97 %). Figure 4 shows the evolution of the value of the activation for each of the event occurrences and the decisions made.



(a) HIS

(b) Sweet-Home

Fig. 4: Example showing changes in levels of activation contexts and selected contexts for the possible locations

V. 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 it is possible to leverage the semantics of the events to gain a higher accuracy as was done with the contact sensors of the room doors for the Sweet-Home corpus. In that case, the knowledge of room transitions was expressed in terms of conditional probabilities and its exploitation was demonstrated to be useful.

The next step is to continue the implementation of the intelligent controller, to implement the voice command recognition, and to test the general suitability of the approach by confronting the system to actual users (elderly and frail people). More challenging tasks will be to make speech recognition robust to environmental noise and to be able to deal with several users.

Acknowledgment

The authors would like to thank the participants who performed the experiments. This work is supported by the Agence Nationale de la Recherche (ANR-09-VERS-011).

References

[1] X. Bian, G. D. Abowd, and J. M. Rehg, “Using sound source localization in a home environment,” in *Third International Conference of Pervasive Computing*, 2005, pp. 19–36.

[2] A. Fleury, M. Vacher, F. Portet, P. Chahuara, and N. Noury, “A multimodal corpus recorded in a health smart home,” in *LREC Workshop Multimodal Corpora and Evaluation*, Matla, 2010, pp. 99–105.

[3] M. Hamill, V. Young, J. Boger, and A. Mihailidis, “Development of an automated speech recognition interface for personal emergency response systems,” *Journal of NeuroEngineering and Rehabilitation*, vol. 6, 2009.

[4] M. Klein, A. Schmidt, and R. Lauer, “Ontology-centred design of an ambient middleware for assisted living: The case of soprano,” in *30th Annual German Conference on Artificial Intelligence (KI 2007)*, 2007.

[5] 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*, vol. 10, no. 1, pp. 92 – 99, 2006.

[6] R. López-Cózar and Z. Callejas, “Multimodal dialogue for ambient intelligence and smart environments,” in *Handbook of Ambient Intelligence and Smart Environments*, H. Nakashima, H. Aghajan, and J. C. Augusto, Eds. Springer US, 2010, pp. 559–579.

[7] S. Moncrieff, S. Venkatesh, and G. A. W. West, “Dynamic privacy in a smart house environment,” in *IEEE Multimedia and Expo*, 2007, pp. 2034–2037.

[8] M. E. Niessen, L. van Maanen, and T. C. Andringa, “Disambiguating sounds through context,” in *Proceedings of the 2008 IEEE International Conference on Semantic Computing*. IEEE Computer Society, 2008, pp. 88–95.

[9] M. Vacher, A. Fleury, F. Portet, J.-F. Serignat, and N. Noury, *Complete Sound and Speech Recognition System for Health Smart Homes: Application to the Recognition of Activities of Daily Living*. Intech Book, 2010, pp. 645 – 673.

[10] M. Vacher, F. Portet, A. Fleury, and N. Noury, “Development of audio sensing technology for ambient assisted living: Applications and challenges,” *International Journal of E-Health and Medical Communications*, vol. 2, no. 1, pp. 35–54, 2011.

[11] C. R. Wren and E. M. Tapia, “Toward scalable activity recognition for sensor networks,” in *Location- and context-awareness*, 2006.