34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

First Steps in Adaptation of an Evidential Network for Data Fusion in the Framework of Medical Remote Monitoring

P.A. Cavalcante, M.A. Sehili, M.Herbin, D. Istrate, F. Blanchard, J. Boudy and B. Dorizzi

Abstract— **This paper presents a medical remote monitoring application which aims at detecting falls. The detection system is based on three modalities: a wearable sensor, infrared sensors and a sound analysis module. The sound analysis is presented briefly. The multimodal fusion is made using the Dempster Schaffer theory through Evidential Network. A first evaluation of the use of data mining techniques in order to extract blindly data representatives is proposed. These representatives are used to continuously increase the system performances. The system is evaluated on a local recorded data base.**

I. INTRODUCTION

Life expectancy is currently increasing in the majority of countries. In 2007, life expectancy was about 79 years (women) and 71 (men) in the world and about 84 years (women) and 77 (men) in France (source: INSEE 2007).

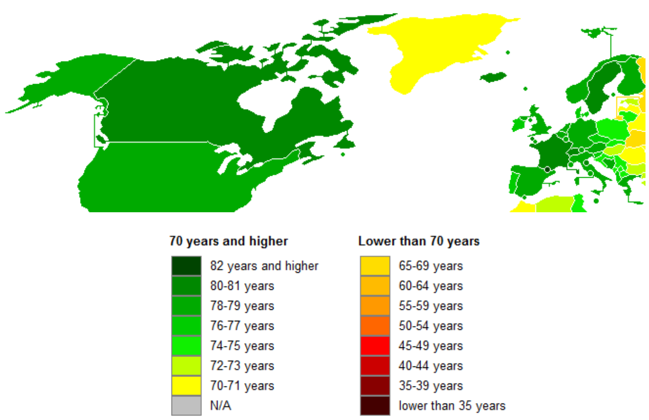

Figure 1 - Life expectancy in the world (ONU, 2007)

In France, 4.5% of men and 8.9% of women aged more than 65 years has had an accident. Among the elderly, 61 % of the accidents occur at home, and 54% take place in the house. In France also 2 million elderly people fall every year. The consequences of these falls are: 10000 deaths, 30-55 % contusions, 3-13 % fractures, dislocations of an articulation, and shock. The total cost is estimated at 1,034 billion € which represent 10 % of the total health costs in France.

The most well known medical remote monitoring systems developed and evaluated up to now [1,2] include distress situation identification and especially the fall detection feature. Different sensors are used: wearable sensors (accelerometers, magnetic sensors, etc.), video analysis, infrared sensors and sound analysis. Generally a multi-sensor fusion is proposed in order to provide more accurate and reliable information. The potential possible benefits of multi-sensor fusion are the redundancy and complementarity of information. The fusion of redundant information can reduce the overall uncertainty. Moreover, the data from multiple heterogeneous sensors of the medical remote systems present uncertainty and lack of confidence [1, 2].

Among multi-sensor fusion techniques, we have Fuzzy logic, Bayesian methods [3] and the Theory of Evidences based on the Dempster-Shafer theory [4], which are commonly used to process and estimate degrees of uncertainty in the fusion process. These theories are based on graphical representations: Bayesian Networks and Evidential Networks (EN) [5].

This paper proposes and evaluates an evidential network to detect fall situations through a heterogeneous multi-sensors fusion. The original aspect of this work consists of the coupling with a data mining algorithm which enables the continuous adaptation of the EN configuration to the real data.

This research is being conducted under the European Project CompanionAble[1] an internationally active group dedicated to carrying out leading-edge research in computer vision and signal processing for human-machine communication, including patient home-care, gesture-based interaction, biometry, video surveillance. The resulting evidential network will be further applied in the French National Project Sweet-Home which aims at supplying a user-friendly (sound/speech and touch) interface to domotics.

II. PROPOSED MEDICAL REMOTE MONITORING PLATFORM

A medical remote monitoring platform was developed at Telecom SudParis elaborated with the close collaboration of ESIGETEL and U558-INSERM. This distress situation detection platform is composed of three modalities: infrared sensors (GARDIEN [7]), a wearable sensor (RFPAT [8]) and sound analysis (ANASON [6]). The platform comprises two simulated rooms architecture filled with the modalities sensors (more details in [9]).

RFPAT system was designed for remote monitoring of vital and actimetric signals recorded on the person. This system is composed of a wearable terminal carried by the patient that can automatically identify distress situations

P. A. Cavalcante, J. Boudy and B; Dorizzi are with TSP, 9 Rue Charles Fourier, 91000 evry (phone: +331 60 764663; e-mail: paulo.cavalcante@it-sudparis.eu).

M. Herbin and F. Blanchard are with CReSTIC, Université de Reims Champagne Ardenne, Chaussée du port - BP541 51012 Châlons-en-Champagne (e-mail: michel.herbin@univ-reims.fr).

D. Istrate, M.A. Sehili are with ESIGETEL, 1 Rue du Port de Valvins, 77210 Avon (e-mail: dan.istrate@esigetel.fr).

[1] *www.companionable.net*

such as falls, cardiac problems (mainly bradycardia and tachycardia) or a person's activity (movements, posture).

The GARDIEN system consists of a fixed network of wired or wireless infrared motion sensors placed within the smart home environment and external to the person. These sensors are activated by body movements which therefore indicate the presence of a person in the area of interest. The person's posture inclination can also be estimated from the combination of two types of infrared sensors, one for horizontal detection field, and the other for vertical.

Sound-based solutions are very interesting, not only due to the scientific advances in the sound recognition and Automatic Speech Recognition (ASR) fields but also because of the good quality, low price and intrusiveness level of today's microphones which would make them acceptable by elderly people and their families.

Obviously, the most interesting class of sound is speech, but in realistic environments there are some kinds of sounds that might be very useful to the detection of abnormal situations like a distress situation (crying, screaming etc.) or to monitor the subject's behavior (doors opening, water running, etc.). In this work we focus on the separation between speech and daily sounds.

Although a Voice Activity Detection (VAD) followed by a speech detection approach could be useful in forcing the ASR system to process only the crucial parts of a stream,, the presence of everyday sounds in the ASR input increases the computation burden and considerably reduces the speech recognition performances. Thus a speech/non-speech classification step becomes very desirable.

III. SOUND ANALYSIS MODALITY

Sound recognition methods have taken advantage of progress in the speaker recognition and speaker verification areas. In this work we apply a combination of Gaussian Mixture Models (GMMs) which are generative models [10] and the kernel method Support Vector Machines (SVMs) [11] to the sound classification problem. This method has actually been successfully used for speaker recognition and verification [12] and it has given better results in comparison to standard systems based on GMM.

Sequence discriminant kernels [13] were proposed to cope with data sequences of variable length. This is a very interesting issue, particularly in sound recognition because we often deal with sound signals with an arbitrary number of frames, a fact that makes the classification of a sequence as a whole difficult. The frame level classification approach is very limited in terms of performances.

The basic idea of the kernel we use for sound classification is to transform a sequence of vectors into one high-dimension vector by adapting a huge GMM Universal Background Model (UBM, usually containing 1024 or 2048 components) using the sequence of vectors and taking the adapted mean vectors to form one super vector used as an input for SVM (Figure 2). The GMM UBM is created using sounds from all classes.

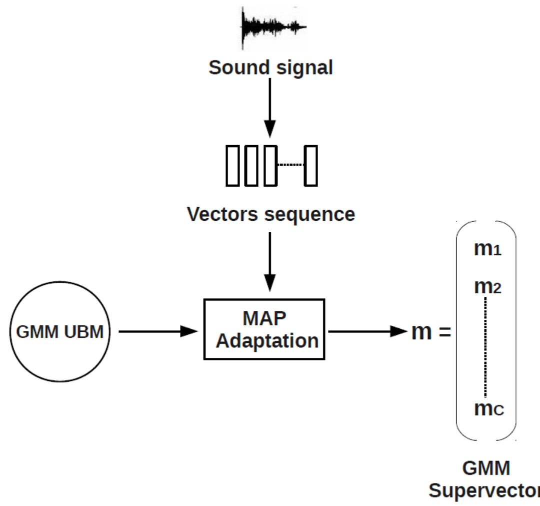

Figure 2 - GMM/SVM approach

IV. EVIDENTIAL NETWORK FOR FALL DETECTION

The proposed evidential network is structured as an acyclic graph and is presented in [9]. Hierarchy and links between nodes create dependencies between the different alarm management modalities (GARDIEN and RFPAT), which makes the fusion process more robust and reliable for abnormal events detections (Figure 3). The sound/speech activity will be added to the network in order to increase performances.

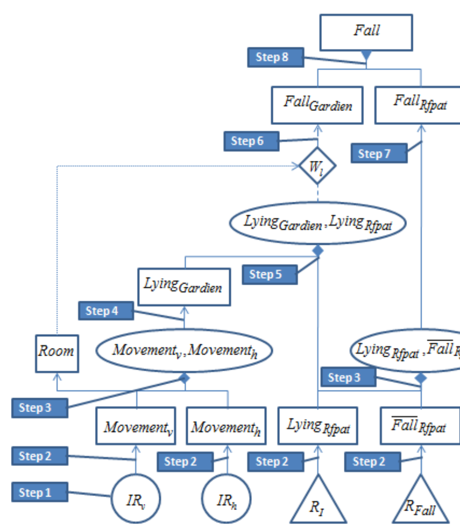

Figure 3 - Evidential Network which fuses infrared sensor and wearable outputs for fall detection

V. DATA ANALYSIS FOR AUTO-ADAPTATION OF THE NETWORK

In the previous section, fall detection is based on an evidential network. In the training step, this approach needs the belief mass and the rules used by the network. Unsupervised data analysis approach gives another way to model the network knowledge. We use this blind approach to adapt the network rules to the fall detection.

First let us describe the data analysis we propose. The sample of multi–sensor data is randomly partitioned into two complementary subsets. The first one is called training subset and the second one validation subset. The variables are rescaled by normalization with the minimum and range. Normalized data become *n*-dimensional vectors in $[\![0,1]\!]^n$ where *n* is the number of variables (i.e. number of sensors). Using the training subset, we compute the correlation coefficients between each normalized variable and the probability of falling (0 no fall, 1 fall). These correlation coefficients are the weights for computing the weighted distances between normalized data. These weighted distances are used to compare and sort data.

The data processing consists of extracting exemplars from the data sample we describe hereafter [14]. The extraction is based on a rank analysis. The weighted distances between an initial datum and the other ones are computed. The distances are relative to the initial datum. These distances are sorted defining the relative ranks. Let N be the number of data. Each datum has N relative ranks. The sum of relative ranks defines the global ranks inside the sample. The lower the global rank of a datum, the more centered the datum inside the whole sample. We consider that the global rank indicates the ability for connecting a datum to the whole sample. When considering the k nearest neighbors (knn) of a datum, this datum is connected to the neighbor having the highest global rank. Then the sample becomes network connecting data. When a datum is connected to itself, then this datum has no better neighbor to connect itself to the whole sample. Such self-connected data is an exemplar. The exemplars are the roots of the trees which form the data network.

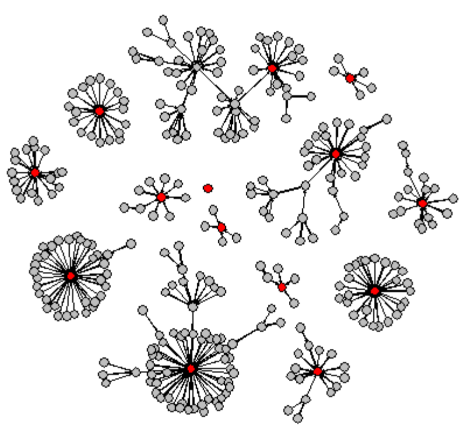

Figure 3 – Blind analysis: example of data network with 14 exemplars (k=10, N=313)

The network and the exemplars depend on the number k of neighbors. If k=1 then each datum is itself its single neighbor. Thus the network has N exemplars (i.e. each datum is itself an exemplar). If k=N then the whole sample is itself the neighborhood of each datum. Then the network has only one exemplar (i.e. the datum with the highest global rank). Therefore the number of detected exemplars decreases from N to 1 where k increases from 1 to N. Moreover this number is always lesser than N+1-k. Let p be the number of detected exemplars. We assume that the higher the difference (N+1-k)-p, the better is the exemplar selection. This criterion is optimized for selecting the optimal number of neighbors. The training set allows us both to compute the weights of the variables (sensors) and to estimate k used to extract the exemplars and to build the network.

Using these estimates, we could apply this blind approach of data analysis on off-line samples (or validation samples) building the data network associated with the samples.

VI. Results

A. Databases

The evaluation of the proposed system was performed on data recorded on the platform of Telecom SudParis. This database consists of simulated scenarios with normal and fall situations. The fall scenarios are composed by 16 classical falls (rather violent) and 17 soft falls (with low acceleration). The ratio between fall and normal scenario is 33/5. The average duration of each scenario is about 10 minutes.

The sound/speech detection is tested on another data base, using five recordings of a duration varying from 14 up to 28 minutes recorded in our laboratory. In each scenario, the subject was asked to read a set of sentences under different configurations. First the subject is close to the microphone and without the presence of noise. Then they move away from the microphone which causes an SNR decrease and finally the noise of a vacuum cleaner is superimposed over the utterances to make the most challenging case.

The UBM model contains 1024 components and was created using speech and sounds from other recording sessions. Some of the used sound classes are: water, door, clapping, vacuum cleaner, radio music, etc. and they are present in the test recordings.

Sound activity is detected using a wavelet based sound detection algorithm and the corresponding signal is framed to compute a sequence of 16 Mel-Frequency Cepstral Coefficients (MFCC) vectors. Then the vectors are used to create the super vector used by SVM.

B. Sound/Speech detection

Out of a total of 213 sentences, 171 sentences were detected and well classified as speech which gives an accuracy of about 80 %. However, it is interesting to mention that most false negative errors were caused by the sound detection step (39 sentences were not detected) against only 3 sentences that were actually detected but mistakenly classified as non-speech.

C. Fall detection

To evaluate the evidential network performance a confusion matrix has been computed on normal and fall situations, as shown in Table 1.

TABLE 1
Confusion matrix of the Evidential Network Data fusion

Confusion matrix	EN fusion	
	Normal	Fall
Normal	5	0
Fall	2	31

In Table 1, "Fall" is a fall detected event and "Normal" is when no fall is detected. The EN fusion has not detected only 2 fall cases which the network is not adapted to: the first case is a soft fall on a sofa; the second case is a soft fall in a bedroom. Concerning the fall on the sofa, the system is not yet adapted: the person falls without impact and ends in a sitting posture / standing. This network needs the lying down posture to detect the fall. Concerning the case of soft falls in the bedroom, the EN fusion allows the detection of this difficult situation for the separated modalities. In this case the context of location "room" is very uncertain and it has a low weight heuristic because the posture "lying down" in the room does not necessarily mean that the person has fallen. It may be a situation where the person is sleeping. This problem can be solved with the introduction of other modalities (video, for example) in the network where we can have more positional accuracy to distinguish; for example if the person is lying because they are in bed or on the ground.. The EN fusion presented promising good performance (sensitive of 93,94%), in particular for soft fall cases, compared to separated modalities.

D. Data Analysis for improving evidential network

The off-line analysis of data samples could allow us to build a data network where the exemplars are the roots of the trees of the network (see section IV). Let us consider fall detection. When an exemplar corresponds to a fall, the whole tree is considered as corresponding to a fall because its root (i.e. the exemplar) is a fall case. Then data belonging to the tree could help to design a new rule for fall detection. This new rule could improve the evidential network by adapting rules and belief mass to these new samples of data.

In the Table we have all 14 specimens for analyzed data. We can observe that the specimen with the most important number of points (841) corresponds to a soft fall (little activity - 4) which has not been detected by the wearable sensor. We have in the table another specimen with 129 points which also correspond to a fall. All falls were simulated in the room corresponding to sensor 1. These specimens with the related points correspond to existing parts of the evidential network.

TABLE 2 - EXEMPLARS OF ANALYZED DATA

N° data	Fall	Position	Activity	Pulse	Horizontal sensor	Vertical sensor
7	0	0	11	56	2	6
51	0	0	13	62	3	5
78	1	1	11	64	1	1
90	0	0	4	64	2	5
30	0	0	3	74	2	6
581	0	0	11	70	2	2
711	0	0	11	76	3	5
841	0	1	4	56	1	1
120	0	0	12	70	3	1
129	0	1	9	66	1	1
157	0	0	12	62	3	4
182	0	0	4	68	3	4
227	0	0	11	70	1	1
237	0	0	3	62	1	1

The data from sensors were recorded during a week and thereafter analyzed by the algorithms without need of human intervention. The exemplars are used to add new analysis parts in the EN.

VII. CONCLUSIONS

This paper presents a medical remote monitoring system enabling the detection of falls using 3 modalities: wearable sensor, infrared sensors and sound analysis. The sound/speech method is presented and evaluated. The data fusion proposed method is based on evidential network. The original aspect of the work consists of coupling this method with a posteriori blind data analysis in order to upgrade the evidential network. Currently the same data is being used to create the evidential network and for data mining analysis. The blind data analysis confirms the configuration of our evidential network.

ACKNOWLEDGMENT

The authors gratefully acknowledge the contribution of European Community's Seventh Framework Program (FP7/2007-2011), CompanionAble Project (grant agreement n. 216487). These results will be further developed in the framework of the French National Project Sweet-Home (ANR VERSO 2009).

REFERENCES

[1] Hyun Lee; Jae Sung Choi; Elmasri, R.; , "Sensor Data Fusion Using DSm Theory for Activity Recognition under Uncertainty in Home-Based Care," Advanced Information Networking and Applications, 2009. AINA '09. International Conference on, May 2009.

[2] Hong, X., Nugent, C., Mulvenna, M., McClean, S., Scotney, B., Devlin, S.: Evidential fusion of sensor data for activity recognition in smart homes, Pervasive and Mobile Computing, Volume 5, Issue 3, Pervasive Health and Wellness Management, June 2009.

[3] Becker, A. and Naïm, P. (1999). Les réseaux bayésiens. Eyrolles, eyrolles edition.

[4] B. Ben Yaghlane, Ph. Smets, K. Mellouli, Directed evidential networks with conditional belief functions. In: Nielsen TD, Zhang NL (eds.) Proceedings of ECSQARU, 2003

[5] Paulo A. Cavalcante A. et al.: Fusion multi-capteurs hétérogène basée sur un Réseau d'Evidence pour la détection de chute. In : ASSISTH conf. 2011.

[6] Istrate, D., Castelli, E., Vacher, M., Besacier, L., Serignat, J. F.: Information Extraction from Sound for Medical Telemonitoring", IEEE Transactions, vol. 10, no. 2, April 2006.

[7] Steenkeste , F., Bocquet, H., Chan, M., Vellas, B.: Remote monitoring system for elders in a geriatric hospital., international conference on aging, Arlington December 1999.

[8] J.L. Baldinger et al.: Tele-surveillance system for patient at home. The MEDEVILLE system. ICCHP, Paris July 2004.

[9] P.A. Cavalcante A., J. Boudy, D. Istrate, H. Medjahed, B. Dorizzi, J. Cesar M. Mota, J. L. Baldinger, T. Guettari and I. Belfeki. "Evidential Network-based Multimodal Fusion for Fall Detection", pHealth 2011

[10] Douglas A. Reynolds and Richard C. Rose, "Robust text-independent speaker identification using Gaussian mixture speaker models," IEEE Transactions on Speech and Audio Processing, pp. 72–80, 1995

[11] Burges Christopher J. C., "A tutorial on support vector machines for pattern recognition," Data Min. Knowl. Discov., pp. 121–167, 1998.

[12] B.G.B. Fauve, D.a Matrouf, N. Scheffer, and J.- F. Bonastre, "State-of-the-art performance in text-independent speaker verification through open-source software," in IEEE Transactions on Audio, Speech, and Language Processing, 2007, vol. 15, pp. 1960–1968.

[13] Tommi Jaakkola and David Haussler, "Exploiting generative models in discriminative classifiers," in In Advances in Neural Information Processing Systems 11.1998, pp. 487–493, MIT Press.

[14] F. Blanchard, P. Vautrot, H. Akdag, and M. Herbin. Data representativeness based on fuzzy set theory. Journal of Uncertain Systems, 4(3):216-228, 2010.