Monitoring of Indoors Human Activities using Mobile Phone Audio Recordings

Prasitthichai Naronglerdrit[1,2], Iosif Mporas[2], Reza Sotudeh[2]
[1]Dept. of Computer Engineering, Faculty of Engineering at Sriracha
Kasetsart University Sriracha Campus
Chonburi 20230, Thailand
[2]School of Engineering and Technology, University of Hertfordshire
Hatfield AL10 9AB, United Kingdom
prasitthichai@eng.src.ku.ac.th, {i.mporas, r.sotudeh}@herts.ac.uk

***Abstract*—In this paper we present a methodology for monitoring of human activities in home using audio recordings captured from mobile phone. Specifically, after estimating a large set of audio features, unsupervised clustering is performed in order to extract feature subspaces. Human activity sound models were trained using different combinations of these subspaces. The best performance 92.46% was achieved using a neural network classifier.**

Keywords—sound recognition; human activity monitoring; classification; clustering.

I. INTRODUCTION

Over the last decade the achievements in areas such as artificial intelligence, social robotics, sensors, statistical signal modeling and machine learning have resulted to the development of autonomous monitoring systems. Monitoring systems are mainly used in applications which promote private/social security, i.e. surveillance systems, or health and well-being. Monitoring systems can roughly be divided into outdoors and indoors, based on the area of application.

Indoors monitoring systems are becoming more and more popular, mainly due to the fact that the ageing population is growing and thus systems that will be able to monitor elders within their homes in order to prevent risks and manage their health status are now essential. Except the ageing population, monitoring systems are also used for the general population in order to provide well-being by analyzing everyday activities and behaviors.

Human activity monitoring is based on data acquisition from a number of sensors, which are processed and analyzed in order to detect specific human activity patterns, such as cooking, eating, walking, showering, flushing etc. The detected human activity events as well as their time and duration log files are further processed in order to understand the affective status of the subject, his/her personal hygiene and everyday habits. Adaptive and personalized behavior models can be developed based on this information which can be on a continuous 24/7 basis. Also, the human activity detectors can be used for detecting hazardous situations, e.g. human body fall, or abnormal situations and subsequently activate an alarm or inform the care-giving person or directly warn the subject him/herself.

For the data acquisition several sensors and modalities have been used. Data acquisition can be performed either using one type of sensors or using multimodal input channels, which are afterwards typically fused on model or decision level.

One of the most popular type of sensors for human activity monitoring is accelerometers [1, 2, 3, 4, 5, 6]. They are used to measure acceleration along a sensitive axis [7] and are quite popular in monitoring human activities related to body movements such as fall detection. When monitoring human activities, the drawback of accelerometer-based monitoring is that they are restricted to human body movements and cannot detect activities with less significant body movement such as singing or watching television.

Other sensors that have been used for human activity monitoring are wearable heart rate [8, 9] and body temperature [10, 11], vibration sensors (geophones) [12] and force sensing resistors embedded in the insole of shoe [13].

Video-based human activity monitoring is the most widely-used modality [14, 15, 16, 17]. Video-based monitoring allows the detection of silhouettes and objects as well as their position and movements. It is a non-intrusive modality for human activity monitoring since, in the general case, it is not wearable. The main disadvantage of video modality is the low performance when the illumination is not high, i.e. during night monitoring. For such cases, although infrared sensors have been proposed [18], audio-based monitoring of human activities is a solution.

Sound recognition has been proposed in several studies for monitoring of indoor human activities. In [19] the authors have used cepstral audio features (MFCCs) and hidden Markov models (HMMs) to recognize activities within a bathroom. In [20] the

authors are using cepstral features and Gaussian mixture models (GMMs) as a classifier to discriminate the sound activities. In [21] a non-Markovian ensemble voting classification methodology is proposed for the recognition of several home activity sounds. The concept of indoor sound activity simulation was introduced in [22]. In [23] the authors presented a sound-source localization methodology for monitoring in-home activities. The advantage of audio-based monitoring is that can operate equally well in low-illumination conditions and can - in the general case - detect an indoors event, even when barriers such as walls interject, which would be a problem in the case of video modality.

Except single-mode sensors several studies have been done using heterogeneous sensors, such as in [24] where audiovisual and medical sensors were used, in [25] where video and laser scanner were deployed, in [26] where video and accelerometers were combined and in [18] where physiological sensors, microphones and infrared sensors were used. Finally, in [27] the fusion of audio information with accelerometers improved the automatic classification accuracy of physical activities.

In this paper we present a methodology for recognizing sounds produced by human activities in home using a mobile phone microphone. The methodology is based on time and frequency domain audio parameterization algorithms and uses subspaces of the multidimensional feature space during the classification step, which are extracted with respect to the discriminative ability of each of the audio parameters.

The remainder of this paper is organized as follows. In Section 2 we present the proposed methodology for indoors home activity recognition using audio features subspaces. In Section 3 we present the experimental setup and in Section 4 the evaluation results are described. Finally, in Section 5 we conclude this work.

II. Sound Recognition of Indoors Human Activity using Audio Features Subspaces

The proposed methodology for sound recognition of indoors human activities is based on the use of a large number of widely-used time and frequency domain based audio parameterization algorithms in order to exploit the different temporal and spectral

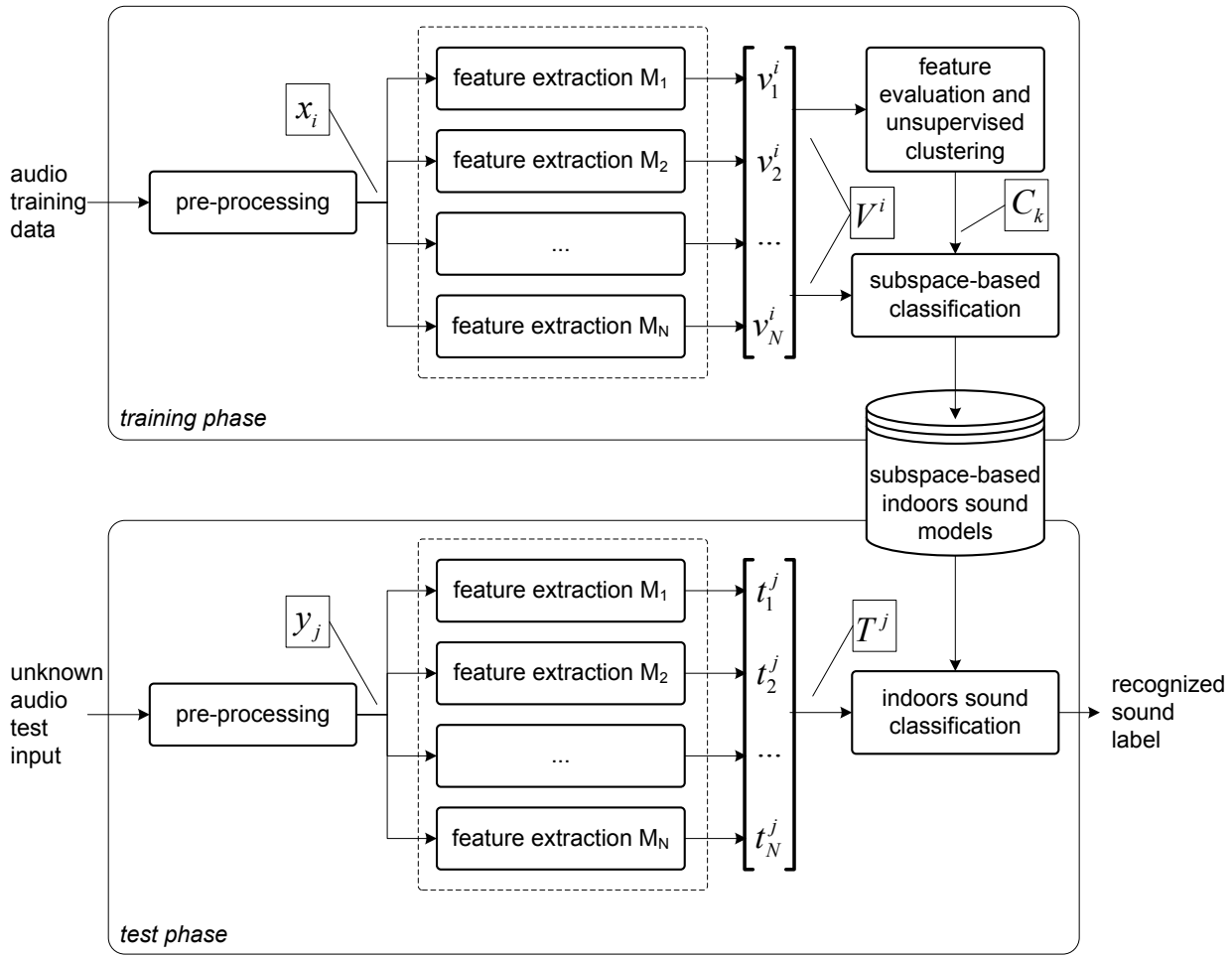

Fig. 1. Block diagram of the sound recognition of indoors human activity using audio feature subspaces.

patterns of sounds produced by human activities in home, as described in Section 1. In particular, during the training phase of the methodology short-time analysis of the audio signal is applied by pre-processing and audio parameterization (feature extraction). Unsupervised clustering of the feature vector space is applied using the decomposed audio signals and feature subspaces are defined. For different groups of subspaces one classification model is built using training data with known sound labels. In the test phase, each unknown audio signal is decomposed to a sequence of feature vectors using the time-domain and frequency-domain feature extraction algorithms used in the training phase and each feature vector is processed by a feature subspace dependent classification model to assign a sound label to each audio frame. The concept of the proposed methodology is illustrated in Figure 1.

As can be seen in Figure 1, during the training phase a training set of audio data with known sound labels (annotated time intervals per human activity sound type) is used. The set of training audio signals is initially preprocessed. Pre-processing consists of constant length frame blocking and constant time-shift between successive frames. The training audio signals are thus decomposed to a sequence of short-time frames x_i, with $1 \le i \le I$. After pre-processing each audio frame, x_i, is used as input to a number of audio parameterization algorithms. Here we consider N time and frequency domain audio parameterization algorithms, M_n, with $1 \le n \le N$. For every audio frame, x_i, each algorithm computes a feature vector, $v_n^i \in \Re^{d_n}$, where d_n is the dimensionality of the features estimated by the n-th audio parameterization algorithm. All estimated feature vectors, v_n^i, $1 \le n \le N$, are concatenated to a single feature vector, $V^i \in \Re^D$, with $D = \sum_{n=1}^{N} d_n$, representing the i-th audio frame. A feature evaluation algorithm is afterwards applied to estimate the discriminative ability of each of the D features. The feature evaluation scores are then processed by an unsupervised clustering algorithm and divide the feature dimensional space into K subspaces, C_k, with $1 \le k \le K$. For different sets of subspaces, sound models are trained using a classification algorithm. The number of feature subspaces is manually selected. During the test phase the unknown audio waveform, y, is pre-processed and parameterized as in training, i.e. is decomposed to audio frames, y_j, with $1 \le j \le J$, and afterwards to a sequence of feature vectors, T^j. Using a subspace-dependent sound classification model, as developed during the training phase, an indoors human activity sound label is assigned to each test audio frame, y_j.

III. Experimental Setup

In this section we describe the audio data used in the current evaluation, the audio features which were estimated and the machine learning algorithms which were used for the clustering and classification steps.

A. Audio Data

The audio data were collected in typical home environment using a conventional mobile phone microphone. All audio data were recorded at sampling frequency equal to 16 kHz with resolution equal to 16 bit per sample. In all recordings the distance of the microphone (i.e. the mobile phone) from the sound source was approximately 1 meter. The types of home activities that were recorded as well as their duration in seconds are tabulated in Table 1. At least 10 different audio recordings for each home activity were done. All audio files were manually annotated using Praat software [28].

TABLE I. List of Recorded Home Activity Sounds and Their Duration in Seconds

Home Activity	Duration (sec)
Background silence	120.54
Door opening/closing	16.50
Door slamming	6.04
Door locking	16.28
Door knocking	15.78
Flushing	232.80
Tap water	340.54
Showering	83.36

Putting object on table	19.60
Switch on/off	6.38
Footsteps	163.42
Spraying	32.06
Vacuuming	237.92
Mixing with spoon	178.52

B. Audio Parameterization

The audio waveforms were decomposed to feature vectors using a wide range of well know and widely used time and frequency domain audio parameterization algorithms. These are: the zero crossing rate (ZCR); the 12 first Mel frequency cepstral coefficients (MFCCs) including the 0-th coefficient; the frame energy (E); 20 linear prediction coding (LPC) coefficients; the harmonics to noise ratio (HNR); the minimum, maximum and mean sample value per frame; the 25, 50 (median) and 75% percentiles of the per frame samples; the 64-bin spectral magnitude; the energy entropy; the spectral entropy, centroid, flux and roll-off; the 12-first chroma coefficients. The total dimensionality of the feature space was D=123.

C. Unsupervised Clustering

The clustering of the feature space was performed on the basis of the discriminative ability of the audio features. We applied feature ranking using the ReliefF algorithm [29], which evaluates the worth of a feature by repeatedly sampling an instance and considering the value of the given feature for the nearest instance of the same and different class. Based on the feature ranking scores we applied expectation-maximization (EM) clustering to 10 clusters.

D. Classification Algorithms

For the development of the home activities sound models we relied on a number of widely used in the area of audio processing classification algorithms. These algorithms are: the naive Bayes algorithm (NB); the k-nearest neighbor classifier (KNN); the C4.5 decision tree (C4.5); the support vector machines (SVM) using the sequential minimal optimization implementation and a radial basis function kernel; the multilayer perceptron neural network with one hidden layer (NN). For the development of the home activities sound models we relied on the WEKA software toolkit implementations [30].

IV. Experimental Results

The home activity sound monitoring methodology presented in Section 2 was evaluated following the experimental setup described in Section 3. In order to evaluate the performance of the proposed methodology we examined the percentage of correctly classified audio frames. The audio dataset presented in Section 3 was modified by randomly selecting subset of the audio frames for the sound types with extensive duration in order to be comparable with most of the other sound types during classification. We followed a ten-fold cross validation setup, in order to avoid overlap between training and test datasets.

During the training phase we applied unsupervised clustering of the feature space to 10 feature subspaces, based on the ReliefF feature evaluation scores. The number of the features per subspace are tabulated in Table 2.

TABLE II. Number of Features per Subspace after Unsupervised Clustering

subspace	**#features**
1	6
2	9
3	7
4	8
5	27
6	4
7	44
8	5

9	7
10	6

The first subspace includes the most discriminative audio features, while the last subspace (10) includes the less discriminative ones. The features which were clustered to belong to the first subspace are: the spectral entropy; the 0-th MFCC coefficient; the HNR; the spectral roll-off; the 3-rd and 1-st MFCC coefficients. The features which were clustered to belong to the second subspace are: the 4-th MFCC coefficient; the ZCR; the 7-th and 2-nd MFCC coefficient; the spectral centroid; the 12-th MFCC coefficient; the 11-th and 6-th chroma coefficients; the 1-st LPC coefficient. The features which were clustered to belong to the second subspace are: the 5-th, 11-th and 6-th MFCC coefficient; the energy entropy; the 2-nd LPC coefficient; the 8-th MFCC coefficient; the minimum frame sample value. These results indicate that the most discriminative audio features are the spectral entropy, the harmonics to noise ratio and the first 3 MFCC coefficients.

In a second step we investigated the performance of the classification algorithms presented in Section 3. The classification accuracy, on audio frame level, for different number of subspace sets and for different classifiers is tabulated in Table 3. The best achieved performance per classification algorithm is shown in bold.

TABLE III. HOME ACTIVITIES SOUND RECOGNITION USING FEATURE SUBSPACE SOUND MODELS. ACCURACIES ARE IN PERCENTAGES

subspace model	**NB**	**KNN**	**C4.5**	**SVM**	**NN**
C_1	66.40	82.96	83.85	71.06	80.98
C_{1-2}	78.32	87.76	87.27	83.60	88.81
C_{1-3}	83.08	89.40	87.64	86.78	90.85
C_{1-4}	**83.30**	89.44	**87.71**	87.64	91.75
C_{1-5}	79.25	89.53	87.15	88.80	92.40
C_{1-6}	75.38	**89.54**	87.07	88.85	**92.46**
C_{1-7}	61.52	89.47	86.80	89.01	90.39
C_{1-8}	59.04	89.46	86.93	**89.03**	88.71
C_{1-9}	56.44	89.47	86.91	89.02	88.55
C_{1-10}	55.01	89.47	86.74	89.02	88.45

As can be seen in Table 3, the best performance was achieved by the neural network, the k-nearest neighbor and the support vector machines algorithm. In detail, the NN model achieved 92.46% accuracy for the subspace model C_{1-6} (i.e. for the first 6 subspace features, which have total dimensionality equal to 6+9+7+8+27+4=61), while the KNN model achieved 89.54% accuracy also for the subspace model C_{1-6}. The SVM model achieved home activities sound recognition accuracy equal to 89.03% for the subspace model C_{1-8} (i.e. for the first 8 subspace features, which have total dimensionality equal to 6+9+7+8+27+4+44+5=110). The rest two classification algorithms NB and C4.5 achieved 83.30% and 87.71% respectively.

In all evaluated classifiers the subspace sound model methodology presented in Section 2, resulted to improvement of the overall performance. Exception is the SVM algorithm, which did not achieve significantly better performance comparing to the full feature space model (C_{1-10}), due to the fact that SVMs circumvent the 'curse of dimensionality' [31]. The presented methodology for modeling of the home activities sounds using feature spaces of lower dimensionality without sacrificing the recognition performance can be used in real-life applications, such as on mobile applications were low computational complexity during the operational (test) phase is a must.

V. CONCLUSION

Among other modalities, monitoring of human activities in home can be performed using sound recognition. The use of sound for human activity monitoring is non-intrusive to the user and allows the monitoring of activities even during night, where typical video monitoring systems are not accurate.

We presented a methodology for the monitoring of indoors activities using audio signal captured by conventional microphone of a mobile phone. The methodology estimated audio feature subspaces by unsupervised clustering of the whole feature space and

models the human activities sounds using combinations of them. The experimental results indicated the validity of the methodology for all evaluated algorithms. The best achieved performance was 92.46%, in terms of audio frame classification accuracy, when using a neural network as classifier. We deem the presented methodology can support monitoring systems in which the event detection is performed on a mobile device and thus low computational complexity is required.

References

[1] O. Politi, I. Mporas, V. Megalooikonomou, "Human Motion Detection in Daily Activity Tasks using Wearable Sensors", In Proc. of the 22nd European Signal Processing Conference, EUSIPCO 2014, pp. 2315 - 2319, Lisbon, Portugal.

[2] D. Naranjo-Hernández, L. M. Roa, J. Reina-Tosina, M. Ángel Estudillo-Valderrama, "SoM: A Smart Sensor for Human Activity Monitoring and Assisted Healthy Ageing", IEEE Transactions on Biomedical Engineering, Vol. 59, no. 11, Nov. 2012.

[3] O. Politi, I. Mporas, V. Megalooikonomou, "Comparative Evaluation of Feature Extraction Methods for Human Motion Detection", Artificial Intelligence Applications and Innovations, Volume 437 of the series IFIP Advances in Information and Communication Technology pp 146-154, AIAI 2014 MHDW.

[4] N.F. Ince, Cheol-Hong Min, A.H. Tewfik, "Integration of Wearable Wireless Sensors and Non-Intrusive Wireless in-Home Monitoring System to Collect and Label the Data from Activities of Daily Living", Proceedings of the 3rd IEEE-EMBS International Summer School and Symposium on Medical Devices and Biosensors, MIT, Boston, USA, Sept.4-6, 2006.

[5] G. Uslu, Ö. Altun, S. Baydere, "A Bayesian approach for indoor human activity monitoring", 2011 11th International Conference on Hybrid Intelligent Systems (HIS), 5-8 Dec. 2011, pp. 324 - 327.

[6] H. Thiruvengada, S. Srinivasan, Aca Gacic, "Design and implementation of an automated human activity monitoring application for wearable devices", 2008 IEEE International Conference on Systems, Man and Cybernetics, 12-15 Oct. 2008, 2252 - 2258.

[7] Subhas Chandra Mukhopadhyay, "Wearable Sensors for Human Activity Monitoring: A Review", IEEE Sensors Journal, Vol. 15, no. 3, March 2015, pp. 1321-1330.

[8] L. Yan, J. Bae, S. Lee, T. Roh, K. Song, and H.-J. Yoo, "A 3.9 mW 25-electrode reconfigured sensor for wearable cardiac monitoring system," IEEE J. Solid-State Circuits, vol. 46, no. 1, pp. 353–364.

[9] N. Ravanshad, H. Rezaee-Dehsorkh, R. Lotfi, and Y. Lian, "A levelcrossing based QRS-detection algorithm for wearable ECG sensors,"IEEE J. Biomed. Health Inform., vol. 18, no. 1, pp. 183–192.

[10] J. Pärkkä, M. Ermes, P. Korpipää, J. Mäntyjärvi, J. Peltola, and I. Korhonen,, "Activity classification using realistic data from wearable sensors," IEEE Trans. Inf. Technol. Biomed., vol. 10, no. 1, pp. 119–128.

[11] V. Leonov, "Thermoelectric energy harvesting of human body heat for wearable sensors," IEEE Sensors J., vol. 13, no. 6, pp. 2284–2291.

[12] I. V. Gabriel, P. Anghelescu, "Vibration monitoring system for human activity detection", International Conference ECAI 2015, Electronics, Computers and Artificial Intelligence, June 2015.

[13] A. Shad, E. Rodriguez-Villegas, "Proof of concept of a shoe based human activity monitor", 34th Annual International Conference of the IEEE EMBS, San Diego, California USA, 2012.

[14] Z. Zhou, W. Dai, J. Eggert, J. T. Giger, J. Keller, M. Rantz and Z. He, "A Real-time System for In-home Activity Monitoring of Elders", 31st Annual International Conference of the IEEE EMBS, Minneapolis, Minnesota, USA, 2009.

[15] S. Mohsen Amiri, M. T. Pourazad, P. Nasiopoulos, V. C. M. Leung, "Non-intrusive human activity monitoring in a smart home environment", 2013 IEEE 15th International Conference on e-Health Networking, Applications & Services (Healthcom), Oct. 2013, pp. 606-610.

[16] N. Zouba, F. Bremond, M. Thonnat, "An Activity Monitoring System for Real Elderly at Home: Validation Study", 2010 Seventh IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 2010, pp. 278-285.

[17] Z. Zhou, Xi Chen, Yu-Ch. Chung, Z. He, T. X. Han and J. M. Keller, "Activity Analysis, Summarization, and Visualization for Indoor Human Activity Monitoring", IEEE Transactions on Cicuits and Systems for Video Techonology, Vol. 18, no. 11, Nov. 2008, pp. 1489-1498.

[18] H. Medjahed, D. Istrate, J. Boudy and B. Dorizzi, "Human Activities of Daily Living Recognition Using Fuzzy Logic For Elderly Home Monitoring", IEEE FUZZ 2009, pp. 2001-2006.

[19] J. Chen, A. H. Kam, J. Zhang, N. Liu, L. Shue, "Bathroom Activity Monitoring Based on Sound", Pervasive Computing, Volume 3468 of the series Lecture Notes in Computer Science pp 47-61.

[20] L. Vuegen, B. Van Den Broeck, P. Karsmakers, H. Van hamme, B. Vanrumste, "Automatic Monitoring of Activities of Daily Living based on Real-life Acoustic Sensor Data: a preliminary study", 4th Workshop on Speech and Language Processing for Assistive Technologies SLPAT 2013, pp. 113–118.

[21] J. A. Stork, L. Spinello, J. Silva, K. O. Arras, "Audio-based human activity recognition using Non-Markovian Ensemble Voting", 2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication, Sept. 2012, pp. 509 - 514.

[22] N. F. Ince, C.-H. Min, A. H. Tewfik, "Integration of Wearable Wireless Sensors and Non-Intrusive Wireless in-Home Monitoring System to Collect and Label the Data from Activities of Daily Living", Proceedings of the 3rd IEEE-EMBS International Summer School and Symposium on Medical Devices and Biosensors, MIT, Boston, USA, Sept.4-6, 2006.

[23] Xuehai Bian, Gregory D. Abowd, James M. Rehg, "Using Sound Source Localization to Monitor and Infer Activities in the Home", Technical Report, Georgia Institute of Technology, 2004.

[24] A. Marsh, C. Biniaris, R. Velentzas, J. Leguay, B. Ravera, M. Lopez-Ramos, E. Robert, "A Multi-Modal Health and Activity Monitoring Framework for Elderly People at Home", Handbook of Digital Homecare, Part of the series Series in Biomedical Engineering pp 287-298.

[25] C. Motamed, R. Lherbier, D. Hamad, "A multi-sensor validation approach for human activity monitoring", 2005 7th International Conference on Information Fusion, 2005, pp. 1548-1553.

[26] L. Tao, T. Burghardt, S. Hannuna, M. Camplani, A. Paiement, D. Aldamen, M. Mirmehdi, I. Craddock, "A Comparative Home Activity Monitoring Study using Visual and Inertial Sensors", 17th International Conference on E-Health Networking, Application and Services (IEEE HealthCom), pp. 644-647.

[27] H. Ketabdar, J. Qureshi, P. Hui, "Motion and audio analysis in mobile devices for remote monitoring of physical activities and user authentication", Journal of Location Based Services, Vol. 5, 2011 - Issue 3-4.

[28] P. Boersma, D. Weenink, "Praat: doing phonetics by computer - Computer program", Version 5.3.51, http://www.praat.org/.

[29] I. Kononenko, "Estimating Attributes: Analysis and Extensions of RELIEF", European Conference on Machine Learning, 1994, pp. 171-182.

[30] I.H. Witten, E. Frank, "Data mining: practical machine learning tools and techniques", 2nd ed, Morgan-Kaufman Series of Data Management Systems, San Francisco: Elsevier, 2005.

[31] C. Burges, "A tutorial on Support Vector Machines for Pattern Recognition", Data Mining and Knowledge Discovery, vol. 2(2), 1998, pp. 121-167.