Università Politecnica delle Marche
Scuola di Dottorato di Ricerca in Scienze dell'Ingegneria
Curriculum in Ingegneria Elettronica, Elettrotecnica e delle Telecomunicazioni

Ambient Intelligence: Computational Audio Processing For Human Fall Detection

Ph.D. Dissertation of:
Diego Droghini

Advisor:
Prof. Francesco Piazza

Coadvisor:
Prof. Roberto Bedini

Curriculum Supervisor:
Prof. Francesco Piazza

XVII edition - new series

Università Politecnica delle Marche
Scuola di Dottorato di Ricerca in Scienze dell'Ingegneria
Curriculum in Ingegneria Elettronica, Elettrotecnica e delle Telecomunicazioni

Ambient Intelligence: Computational Audio Processing For Human Fall Detection

Ph.D. Dissertation of:
Diego Droghini

Advisor:
Prof. Francesco Piazza

Coadvisor:
Prof. Roberto Bedini

Curriculum Supervisor:
Prof. Francesco Piazza

XVII edition - new series

Università Politecnica delle Marche
Scuola di Dottorato di Ricerca in Scienze dell'Ingegneria
Facoltà di Ingegneria
Via Brecce Bianche – 60131 Ancona (AN), Italy

alla mia famiglia

Abstract

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Outline, obiettivi, controbuti

The decreasing birth rate [1] and the contemporary increase of the life expectancy at birth [2] in the majority of industrialized countries have been generating new challenges in the assistance of the elderly. The scientific community, companies and governments are trying to face them by investing in the development of efficient healthcare systems and solutions. The direction taken goes towards the development of smart home capable of taking care of the inhabitants by supporting and monitoring them in their daily actions [3, 4]. Since falls are one of the main cause of death for the elderly [5], several efforts have been devoted to the development of algorithms for automatically detecting these events.

1.1 Fall Detection Systems

The continuous and unprecedented growth rate of the elderly world population is one of the primary aspects of concern for society and governments. Nowadays about 8.5% of people in the world are more than 65 years old [6, 2]. Although the average life of the world population is getting longer, elderly people may not necessarily live a healthier life. It is enough to say that 37.5 million falls require medical interventions and more than 600 thousand are cause of death every year worldwide. In particular, the population segment most affected by this problem is composed of elderly over 65 years that, with the growing mobility of the population, are more frequently left alone in their homes without aid in the case of need. Moreover, since falls are the leading cause of death and hospitalizations for older adults, this phenomenon leads to a substantial increase in the cost of healthcare [7, 5]. It is not surprising, thus, that the research community is encouraged, even by governments, to find reliable and performing solutions to minimize the damage caused by the human falls problem. This is also confirmed by the presence in the literature of several reviews dedicated to this specific topic [5, 8, 9, 10, 11, 12]. In fact, in the past few years, a variety of systems have been presented. One way to divide the methodologies for approaching the falls detection problem is based on the placement of the sensing devices [5]. The main categories are wearable, vision and environmental, with each category presenting their own advantages and disadvantages. Wearable systems do not suffer from ambient condition, but people may forget to wear them and they are not operational during the charging time, thus, some people may consider them annoying. Furthermore, a device must be installed on each person to be monitored. An environmental sensor may be used to avoid this kind of problems, but with other limitations. Vision systems, although they are actually environmental sensors, deserve a dedicated category because of many systems proposed in the literature based on this type of sensors [5]. This category includes several types of sensors like, e.g., cameras for which the major limitations are field-of-view constraints, lighting condition, positioning of multiple cameras and lack of privacy. The ambient category includes several types of sensors. For example, radar doppler based systems used in [13] raise fewer privacy concerns, but they suffer from reflection and blind spots. In particular, for a data-driven system, another aspect that should not be underestimated is the need for a re-training when changing the environment to be monitored or even just some of its components such as the arrangement of furniture as happens in [14]. All this implies that there is no optimal choice, which is instead, a compromise that depends on the type of environment that is monitored as well as on personal sensitivity of the subjects under monitoring. Going into more detail, another significant distinction between falls detection systems can be made based on the type and amount of data used for the algorithm development [8]. In fact, the problem can be approached either as supervised or unsupervised based on the availability of data in the hands of the researchers as well as their goals. Most state-of-the-art works tackle the problem under fully supervised conditions assuming they have enough data for falls. Almost all of these falls are simulated with professional mannequins [15, 16] or by people with adequate protections [17, 18] that however may not correctly emulate an actual fall. Although this approach leads to more accurate results, there is no guarantee that it will generalize well in real situations. Other researchers opt for approaches based on outlier/anomaly detection [19, 20, 21] because of the plentiful availability of data that can represent normal activity. However, it is challenging to define what "normal activities" are for such approaches, and the risk is to raise several false alarms. Perhaps the situation that most closely approximates reality is a hybrid between the previous ones, in which a large amount of data representing the normality are easily available, with just a few samples of real human fall (*RHF*) and eventually some related synthetic or simulated data. In these situations, supervised approaches that suffer from strong data imbalance have to apply subsampling [22] or weighting [8] techniques to mitigate this effect. Thus, the need to find an effective way to exploit the few available falls data is evident.

1.2 State-Of-The-Art

Review dei sistemi per la fall detection basati sui vari tipi di sensori accelerometers, vision, ambient. Per gli ambient paricolare enfasi sugli approcci basati su aduio.

As aforementioned, fall detection approaches can be divided based on their sensing technology, in particular if they employ wearable or ambient sensors. Regarding the first ones, the most common choice is to employ accelerometers. The algorithms proposed in [23] and [24] detect a fall by verifying if the acceleration signals exceed a certain threshold. In contrast, in [25] the authors implemented several machine learning techniques and studied their classification performance. For the experiments, a fall events dataset has been developed using six sensor units with three-axis accelerometers and worn by 14 persons who simulated falls from different angles. The best performing classifier resulted the *k*-Nearest Neighbour classifier. [26] employed radio tags worn on the user's chest, waist, and ankles, and an optional three-axial accelerometer worn on the chest. The algorithms performs a basic activity recognition, distinguishing from walking, standing, sitting, sitting on the ground, lying down, the process of sitting or lying down, the process of standing up, and falling. The actual fall is detected combining the results of two classifiers, an SVM and a decision tree, and hand-crafted rules. The authors performed the experiments on a laboratory scenario and reported accuracies of 100% combining radio tags and the accelerometer.

Differently from wearable sensors, the physical quantities captured by ambient sensors are more heterogeneous. Generally, fall detectors are based on vibration, video or acoustic sensors, sometimes in combination with presence detectors. In [27] the fall detector is based on a floor vibration sensor and the algorithm detects a fall when the vibration pattern matches the one of a human fall. The authors do not give further details on the algorithm and report 100% sensitivity and specificity on tests conducted on a dummy falls dataset. Yazar and colleagues [28] employ both passive infrared (PIR) sensors and floor vibration sensors. PIRs are employed to reduce false alarms, i.e., by detecting if a person is present in the region of interest. Single-tree complex wavelet transform features are extracted from the vibration signal and classified as fall or non-fall. In their dataset, the non-fall classes are represented by human (walking or running and sitting) or non-human activities (door slamming and a book falling). Three different classifiers have been compared: Euclidean distance, Mahalanobis distance and SVM, with the latter resulting in the most performing one, since it is able to classify human falls without errors regardless the employment of PIR sensors.

Regarding approaches based on audio signals, a common solution is to install several microphones in the building, usually on the ceiling or near the walls. Indeed, also single-microphone approaches exist but they are much less robust to environmental noise, thus resulting in poor performance. For example, in [29], the authors employ a single far field microphone and they model audio segments by means of perceptual linear predictive (PLP) coefficients and GMM supervectors. An SVM with a kernel based on the Kullback-Leibler divergence, then, classifies the segment as being a fall or noise. For this purpose, nine classes of noise have been considered. In the experiments, the algorithm achieves an F_1-Measure of 67% in the classification task, and an accuracy equal to 64% in the detection task.

The difficulty in using a single microphone drove the scientific community to employ multi-channel algorithms. In [19] the authors present an unsupervised algorithm based on two microphones. The algorithm comprises a source separation and localization block to reduce the impact of background noise. Then, a one class Support Vector Machine is trained on MFCCs of non-fall events only. The SVM is then applied to distinguish normal sound events (i.e., sounds originating from normal activities) from abnormal ones (i.e., falls sounds). The authors validated the algorithm using simulated falls of persons only in presence of a television that produced the interfering sound. The results in terms of Area Under Curve are 0.9928 without interference and 0.9738 with 75% interference. The work by Li and colleagues [17] employs a circular microphone array to firstly determine the position of the sound source, and then to enhance the signal by applying a beamformer. The height of the sound source is used as first filter to discriminate falls from non falls. If the sound originates from a source positioned on the ground, MFCC features are extracted and a k-Nearest Neighbour classifier is employed to detect persons' falls. The algorithm has been tested on a dataset composed of 120 simulated fall sounds and 120 non-fall sounds recorded in different acoustic conditions. In presence of background noise and TV interference, the resulting AUC was equal to 0.989 (accuracy 95%) on clean conditions and 0.932 at 10 dB SNR (accuracy 89%).

An approach to improve the performance of fall detection systems is to combine the information coming from different sensors. The approach proposed by Zigel *et al.* [16] is based on a combination of sound and vibration sensors attached to the floor with a adhesive tape. The algorithm employs energy features extracted from the vibration signal to detect the fall event. Then, the event is classified as fall or non-fall with naive Bayes classifier employing features from both the vibration and sound signals. The experiments were conducted on a dataset containing falls of the "Rescue Randy" human mimicking doll and four objects, and the resulting sensitivity and specificity were respectively 97.5%

and 98.6%. [30] fuse the information coming from a Doppler sensor and motion sensors and classify falls with an SVM. The authors report an AUC equal to 0.98 with Doppler sensor only, and a further reduction of false alarms by 63% employing motion sensors information. Motion, sound and video signals are employed in [31]. Signals are captured both from environment sensors and from body sensors. A fall is detected by analysing sounds and motion information, while visual and motion behaviour indicates the severity of the fall. The work by Toreyin and colleagues [32] combines PIRs, microphones and vibration sensors. Signals are processed to extract features in the wavelet domain and and HMM classifier is then employed to detect falls. The authors showed the using PIR signals 100% accuracy can be obtained.

1.2.1 Problem Statement

Chapter 2

Background

In recent years, the IoT revolution has led to the creation of enormous amounts of data. The use of intelligent devices that can interface with cloud computing systems or perform complex calculations directly on board, in homes and cities, has allowed the affirmation of data-driven algorithms compared to other methodologies used so far. In fact, these approaches try to emulate the functioning of the human mind, enabling computers to perform tasks that are unthinkable until now. In this chapter are resumed the data-driven and machine learning algorithms used for developing the proposed methodologies for fall classification systems.

2.1 Support Vector Machines

Support Vector Machines (SVM) [33] are one of the most popular classification algorithms and are well known for their strong theoretical foundations, generalization performance and ability to handle high dimensional data. This section presents an overview of support vector machine, starting with linear SVMs, followed by their extension to the nonlinear case and finally the One-Class SVM for novelty detection.

Linear Support Vector Machines In the binary classification setting, let $((x_1, y_1) \dots (x_n, y_n))$ be the training dataset where $x_i \in \Re^n$ are the n-dimensional feature vectors representing the instances (i.e. observations) and $y_i \in \{-1, +1\}$ be the labels of the instances. Support vector learning is the problem of finding a separating hyperplane that separates the positive examples (labeled +1) from the negative examples (labeled -1) with the largest margin:

$$f(\vec{w}) = \text{sign}(\vec{w}^T \cdot \vec{x} + b), \tag{2.1}$$

where a value of -1 indicates one class, and a value of $+1$ the other class. In the simpler linearly separable problem, the margin of the hyperplane is defined as the shortest distance between the positive and negative instances that are

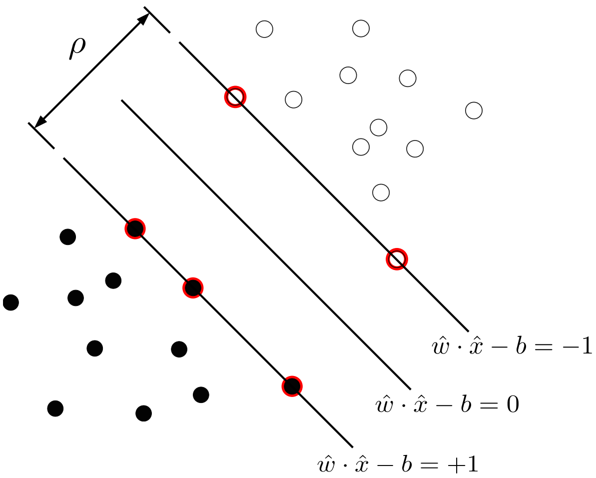


Figure 2.1: A hyperplane separating two classes with the maximum margin. The red highlighted points are the support vectors.

closest to the hyperplane. The intuition behind searching for the hyperplane with a large margin is that a hyperplane with the largest margin should be more resistant to noise than a hyperplane with a smaller margin.

Formally, suppose that all the data satisfy the constraints

$$\vec{w} \cdot \vec{x}_i + b \geq +1 \text{ for } y_i = +1, \tag{2.2}$$

$$\vec{w} \cdot \vec{x}_i + b \leq +1 \text{ for } y_i = -1, \tag{2.3}$$

where $\vec{w}$ is the normal to the hyperplane, $\frac{|b|}{\|\vec{w}\|}$ is the perpendicular distance from the hyperplane to the origin, and $\|\vec{w}\|$ is the Euclidean norm of $\vec{w}$. These two constraints can be expressed in compact form as:

$$y_i(\vec{w} \cdot \vec{x}_i + b) \geq 1. \tag{2.4}$$

The *canonical hyperplane* is the hyperplane that separates the data and has maximal margin. The margin ρ can be computed as the distance between the two canonical hyperplanes:

$$\rho = \frac{1-b}{\|\vec{w}\|} - \frac{-1-b}{\|\vec{w}\|} = \frac{2}{\|\vec{w}\|} \tag{2.5}$$

Thus, we need to solve an optimisation problem, finding the hyperplane that maximises the margin and ensures the classes are separable

$$\min_{\vec{w}_i, b} \frac{1}{2}\|\vec{w}\|^2 \text{ subject to } y_i(\vec{w} \cdot \vec{x}_i + b) \geq 1. \tag{2.6}$$

The problem can be expressed in the Lagrangian formulation:

$$\mathcal{L}(\vec{w}, b, \lambda) = \frac{1}{2}\|\vec{w}\|^2 + \sum_{i=1}^{m} \lambda_i (1 - y_i(\vec{w} \cdot \vec{x}_i + b)) \tag{2.7}$$

with Lagrange multipliers $\lambda_i \geq 0$ for each constraint in 2.6. The objective is then to minimize 2.7 with respect to $\vec{w}$ and b and simultaneously require that the derivatives of $\mathcal{L}(\vec{w}, b, \lambda)$ with respect to all the λ vanish. The advantage is twofold: the training vectors only appear as a scalar product among the vectors, and the constraints are easier to manage.

With the formulation presented above, the SVM fails in some situation. In fact, there is no solution if samples can not be separated by a hyperplane. Moreover, although data are linearly separable the SVM may overfit to some outlier compromising system performance. For dealing with this type of problem, has been developed the soft margin SVM [33] which allows data points to lie within the margins. Introducing *slack variables* ξ_i into the constraints and penalize them in objective, the new problem becomes

$$\min_{\vec{w}_i, b, \vec{\xi}} \frac{1}{2}\|\vec{w}\|^2 + C\sum_{i=1}^{m} \xi_i \tag{2.8}$$

$$\text{subject to } y_i(\vec{w} \cdot \vec{x}_i + b) \geq 1 - \xi_i \text{ and } \xi_i \geq 0 \text{ for } i = 1 \cdots m.$$

The cost coefficient C > 0 is a hyper-parameter that specifies the misclassification penalty and is tuned by the user based on the classification task and dataset characteristics.

Non-Linear Support Vector Machines A way to solve the problem when data are not linearly separable, is to map the data on to a higher dimensional space and then to use a linear classifier in the higher dimensional space. This methods is referred to as "the kernel trick " that exploit the fact that the training data appears as a dot product between vectors in the Lagrangian formulation to from non-linear decision boundaries. Suppose to use a transformation $\Phi : \vec{x} \rightarrow \phi(\vec{x})$ to map every data sample into higher dimensional space, the dot product becomes $\phi(\vec{x}_i)^T\phi(\vec{x}_j)$. By the use of a kernel function

$$K(\vec{x}_i, \vec{x}_j) = \langle \phi(\vec{x}_i), \phi(\vec{x}_j) \rangle, \tag{2.9}$$

it is possible to compute the separating hyperplane without explicitly carrying out the mapping into feature space. The classifier become:

$$f(\vec{x}) = \text{sign}(\sum_i \lambda_i y_i K(\vec{x}_i, \vec{x}_j) + b) \tag{2.10}$$

The most popular kernel functions are:

- Linear Kernel:
$$K(\vec{x}_i, \vec{x}_j) = \langle \vec{x}_i, \vec{x}_j \rangle \tag{2.11}$$

- Polynomial Kernel:
$$K(\vec{x}_i, \vec{x}_j) = (\langle \vec{x}_i, \vec{x}_j \rangle)^d \tag{2.12}$$

- Sigmoid Kernel:
$$K(\vec{x}_i, \vec{x}_j) = tanh(\gamma \langle \vec{x}_i, \vec{x}_j \rangle - \theta) \tag{2.13}$$

- RBF Kernel:
$$K(\vec{x}_i, \vec{x}_j) = \exp(-\frac{\|\vec{x}_i - \vec{x}_j\|}{2\sigma^2}) \tag{2.14}$$

Up to now the SVM algorithm for binary classification has been described. This algorithm can be extended to the multi-class case using the "one vs all" technique [34].

2.1.1 One-Class Support Vector Machines

One-Class SVM (OCSVM) proposed by Schölkopf et al. [35] is the extension of the support vector machine to the case of unlabeled data that makes them useful for novelty detection problems. In the OCSVM, a new parameter ν that controls the trade-off between maximizing the distance of the hyperplane from the origin and the number of data points contained by the hyperplane has been introduced. To separate the data from the origin, the following quadratic program has to be solved:

$$\min_{\vec{w}_i, \vec{\xi}, \rho} \frac{1}{2}\|\vec{w}\|^2 + \frac{1}{\nu l}\sum_{i=1}^{m} \xi_i - \rho \tag{2.15}$$

$$\text{subject to} \quad (\vec{w} \cdot \phi(\vec{x}_i)) \geq \rho - \xi_i \text{ and } \xi_i \geq 0 \text{ for } i = 1 \cdots m.$$

In fact, One-Class SVM consists in a discriminant function that takes the value $+1$ in a small region that captures the majority of the data points of a set and -1 outside that region [36]. The discriminant function has the following expression:

$$f(\mathbf{x}) = \text{sgn}\left(\sum_i \alpha_i \cdot k(\mathbf{x}_i, \mathbf{x}) - \rho\right), \tag{2.16}$$

where $\vec{x}_i$ denotes the i-th support vector. The position of the hyperplane, thus, defines the region that represents normal data points. For each point $\mathbf{x}$ that lies outside this region, the function $f(\vec{x})$ takes the value -1, whereas for point inside the region, it takes the value +1. The terms λ_i can be found by solving the solution to the dual problem:

The terms λ_i can be found by solving the solution to the dual problem:

$$\min_{\lambda} \frac{1}{2} \sum_{ij} K(\vec{x_i}, \vec{x_j}) \tag{2.17}$$

$$\text{subject to} \quad 0 \leq \lambda_i \leq \frac{1}{\nu l} \quad \text{and} \quad \sum_i \lambda_i = 1,$$

where λ_i is a Lagrange multiplier and l is the number of points in the training dataset. The term $\nu \in (0, 1]$ is an hyperparameter of the algorithm that is determined on a validation set.

The offset ρ can be obtained from the Karush-Kuhn-Tucker (KKT) condition with the expression [37]:

$$\rho = \sum_j \lambda_i k(\vec{x}_j, \vec{x}_i), \tag{2.18}$$

which is satisfied for any λ_i that is not at the upper or lower bound.

2.2 Gaussian Mixture Model

https://pdfs.semanticscholar.org/734b/07b53c23f74a3b004d7fe341ae4fce462fc6.pdf A Gaussian Mixture Model (GMM) is a parametric probability density function represented as a weighted sum of Gaussian component densities. Generally, GMMs are used as a parametric model of the probability distribution of some features. To estimate the parameter of GMM the algorithm Expectation-Maximization (EM) algorithm or Maximum A Posteriori (MAP) are used starting from a well-trained prior model usually named Universal Background Model (UBM). A Gaussian mixture model is a weighted sum of M component Gaussian densities as given by the equation

$$p(\vec{x, \lambda}) = \sum_{i=1}^{M} w_i g(\vec{x}|\vec{\mu}_i, \vec{\Sigma}_i) \tag{2.19}$$

where $\vec{x}$ is a D-dimensional features vector, $g(\vec{x}|\vec{\mu}_i), \vec{\Sigma}_i)$ are the components of the mixture and w_i are the weight of each component. Each component of the mixture is a D-variate Gaussian density function expressed as

$$a \tag{2.20}$$

2.3 K-Nearest Neighbor

2.4 Deep Neural Network

2.4.1 Convolutional Neural Network

2.4.2 Autoencoder

Chapter 3

Dataset

The importance of using public data sets for algorithm evaluation is very important. Only in this way can a direct comparison be made between the different approaches to determine which of these is actually the best. There are publicly available datasets for the fall detection task, the majority of them are all related to wearable or vision sensors and often include both types [38, 39, 40, 41, 42]. Since in this work, we face the problem fall detection from an audio perspective, only one dataset containing audio recording has been found [43]. However, the audio files available in the dataset are suitable for speech recognition related works rather than sound event detection. In fact, only the utterance of short sentences or interjections of the actors involved during the human falls recordings have been annotated. As in this work, several data-driven approaches for pattern recognition produced by the sound generated by the human fall are presented, the dataset [43] result useless. Given the lack of available audio data sets, we have created a suitable audio dataset in order to assess the proposed approaches. This choice was also forced by the fact that in these works an innovative acoustic sensor was explicitly developed for the fall detection and described in Section 3.1 has been used. In this chapter, the instrumentation, the procedure for recording the audio corpus and its composition are described.

3.1 The floor acoustic sensor

The floor acoustic sensor (FAS) is composed of a resonant enclosure and a microphone located inside it (Figure 3.1) [44]. At the bottom of the enclosure, a membrane is in direct contact with the floor and guarantees the acoustic coupling with the surface. The inner container accommodates the microphone and is where the acoustic resonance phenomenon takes place. It can be covered by a layer of acoustic isolation material and it is enclosed by the outer container that further reduces the intensity of the acoustic waves that propagate through air. The enclosure has been manufactured in Polylactic Acid with a 3D printer, its diameter is 16.5 cm and its height 5.5 cm.

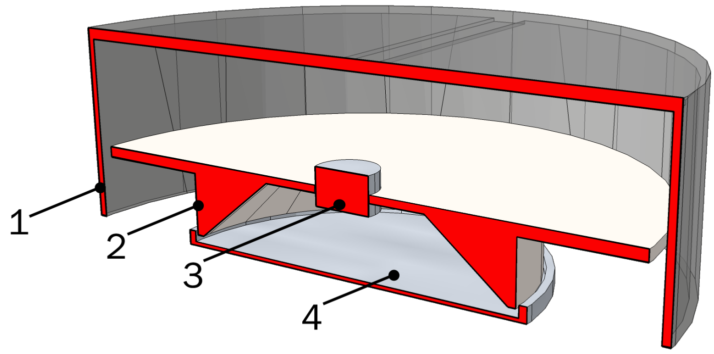


Figure 3.1: The floor acoustic sensor: conceptual scheme. 1 - The outer container. 2 - The inner container. 3 - The microphone slot. 4 - The membrane touching the floor.

Regarding the microphone, an AKG C 400 BL[1] has been inserted in the enclosure. The outer case of the microphone has been removed to extract the capsule that has then been inserted in the sensor enclosure. The AKG C 400 BL is characterized by an hypercardiod directivity pattern, thus it has been oriented so that the maximum gain is towards floor.

3.2 The fall events dataset: A3Fall-v1.0

The performance of the floor acoustic sensor has been evaluated on a corpus of audio events corresponding to falls of several objects recorded in different conditions[2]. The dataset has been specifically created by the authors and it will be presented in this section.

3.2.1 The recording setup

Fall events have been recorded in 3 different rooms with the following characteristics:

- The first, is a rectangular room, hereafter named R0, measuring about $7\,\mathrm{m} \times 2\,\mathrm{m}$ (Figure 3.5). The room is particularly suitable for the propagation of acoustic waves through the floor since it is obtained from a cantilever beam. In addition, the considerable distance of the supporting pillars facilitates the transmission of a fall vibrations through the floor.

[1] http://www.akg.com/pro/p/c400-bl
[2] The dataset is available at the following URL: `http://www.a3lab.dii.univpm.it/research/fasdataset`

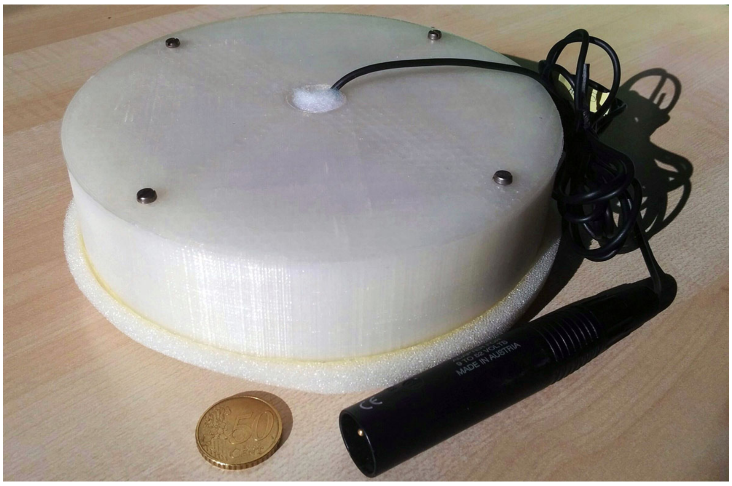


Figure 3.2: A picture of the floor acoustic sensor used during the recordings.

- the second location for the recording was the university auditorium room (R1) in which the flooring is composed of fitted carpet. This makes it particularly suitable for evaluating system performance on surfaces with acoustical behavior that can mitigate the impact sound transmitted through the floor and in the air; all the recordings were performed near the auditorium stage in an area of 8×3 m.

- a recording studio (R2) was selected as the third location for its particular characteristics. Here, it was possible to make the acquisitions by placing the sensors in the live room while the audio events were performed in the control room. In particular, the sensors were positioned immediately behind the soundproof wall with the window overlooking the live room. The size of the live room is 5×7 m, while the size of the control room is 3×8 m.

The recording equipment comprises the floor sensor, a linear array of three aerial microphones (the same AKG 400 BL included in the floor sensor) and a Presonus AudioBox 44VSL sound card connected to a laptop. The microphones of the array are separated by 4 cm and positioned on a table 80 cm high. Signals were sampled at 44.1 kHz with a resolution of 32 bits. Levels were calibrated to assure the maximum dynamic range at the smallest distance.

3.2.2 Description

In Table 3.1 the composition of the dataset is summarized. For the R0, the dataset comprises recordings of fall events related to everyday objects and to

Table 3.1: Composition of the A3Fall-v2.0 dataset.

Class	**R0**	**R1**	**R2**
	Nr. of occurrences		
Basket	64	40	40
Fork	64	40	40
Ball	64	40	40
Book	64	40	40
Bag	64	30	40
Chair	96	40	40
Table	0	40	40
Guitar Slide	0	40	40
Nipper	0	40	40
Keys	0	40	40
Hook	0	40	40
Coat Hook	0	40	40
Manikin Doll	44	0	0
Human Fall	0	40	40
	Total length (s)		
Background	2530	9055	5550

a human mimicking doll. The objects were chosen according to the recent literature on the topic [27] and are the following: a ball, a metal basket, a book, a metal fork, a plastic chair, and a bag (Figure 3.3). Objects have been dropped at four distances from the sensors, i.e., 1 m, 2 m, 4 m, and 6 m, and with various angles in order to reproduce realistically different fall patterns. With the exception of the chair and the basket, which have been overturned from their natural position, half of the falls has been performed at a height of 0.5 m and the other half at 1 m. For each object and for each distance, 16 fall events have been performed for a total of 64 events per object. Instead, the chair has been overturned 8 times for each side of fall (back, front, side) and for each distance, thus obtaining a total of 96 events.

Human falls have been simulated by employing the “Rescue Randy” doll[3], a professional equipment employed in water rescues. It weights 75 kg, it is 1.85 m high, and it is equipped with articulated joints. The doll is made of vinyl and its weight is distributed according to the human weight distribution chart. The doll has been dropped from upright position and from a chair, both forward and backward, for a total of 44 events (Figure 3.4). Differently from the everyday objects, the distribution of the fall events with the distance is not uniform: 10 events have been performed from 2 m, 18 from 4 m (7 of which from the chair),

[3]http://www.simulaids.com/1475.htm

Figure 3.3: Objects employed for creating the fall events dataset.

and 16 from 6 m (6 of which from the chair).

Moreover, several backgrounds sounds has been added to the dataset. Normal activities sounds have been recorded while persons were performing common actions, such as walking, talking, and dragging chairs. Three musical tracks have been played from a loudspeaker and acquired back with the FAS. The first track contained classical music[4], while the second[5] and the third[6] rock music. Musical tracks and normal activities sounds have been divided in segments whose lengths have mean and standard deviation estimated from instances of fall events. In addition, they have been employed alone and to create noisy versions of human and object falls occurrences in order to assess the algorithm in presence of interferences.

In R2 and R1 other every-days objects,in addition to those used in R0, have been recorded for a total of 12 different object fall classes and 1420 instances. While the manikin doll has been used only in R0, in R1 and R2 a total 80 human falls have been performed by 4 people. These falls were performed in different ways: forward, backward and on the side, trying to use the arms to cushion the fall and without any protections. As in R0, also in R2 and R2 all events were performed from 1, 2, 4 and 6 m away from the FAS.

As shown in Table 3.1 background noises have been recorded also in R1 and R2 rooms rooms, which include: human activities noise as, i.e., footsteps, human and phone conversation, dragging objects and so on; classic, rock and pop music played from loudspeakers; TV shows like newscast and satiric.

Since the data relating to rooms R1 and R2 have been collected at different times to those of room R0, in the following chapters, for each proposed ap-

[4] W. A. Mozart, "Piano trio in C major"
[5] Led Zeppelin, "Dazed and confused"
[6] Led Zeppelin, "When the levee breaks"

(a) Fall from upright position. (b) Fall from the chair.

Figure 3.4: Falls of the "Rescue Randy" doll from upright position (a) and from the chair (b).

proach, it will be specified which subset of the total dataset has been used as well as the usage of the noisy version of falls events.

3.2.3 Signal analysis

The signal related to the same fall event acquired with the floor sensor and with the aerial microphone exhibits different spectral characteristics. In this section and in depth analysis of the audio signals acquired in the R0 room is presented. Figure 3.6 shows the spectrograms of a doll fall acquired with the floor sensor (above) and with the aerial microphone (below) in the clean acoustic condition. Observing the figures, it can be noticed that the aerial microphone is more sensitive to high frequencies, in particular to the ones above 1.5 kHz. On the contrary, the majority of the energy of the signal acquired with the floor sensor concentrates below 1 kHz.

This is even more evident by plotting the values of the mel coefficients (Figure 3.7): the first and second mel channels of the FAS, corresponding to the frequency bands 0–128.10 Hz and 61.30-200.60 Hz, are higher respect to the aerial microphone. Channels 3 to 7, respectively corresponding to bands 128.10–279.50 Hz and 458.70–670.70 Hz, are almost equivalent, while from channel 8 (560.30–790.80 Hz) to 29 (6654.60–8000.00 kHz) the aerial microphone mels are greater.

The analysis of noisy signals highlights the different behaviour of the floor sensor respect to the aerial microphone in presence of external interferences.

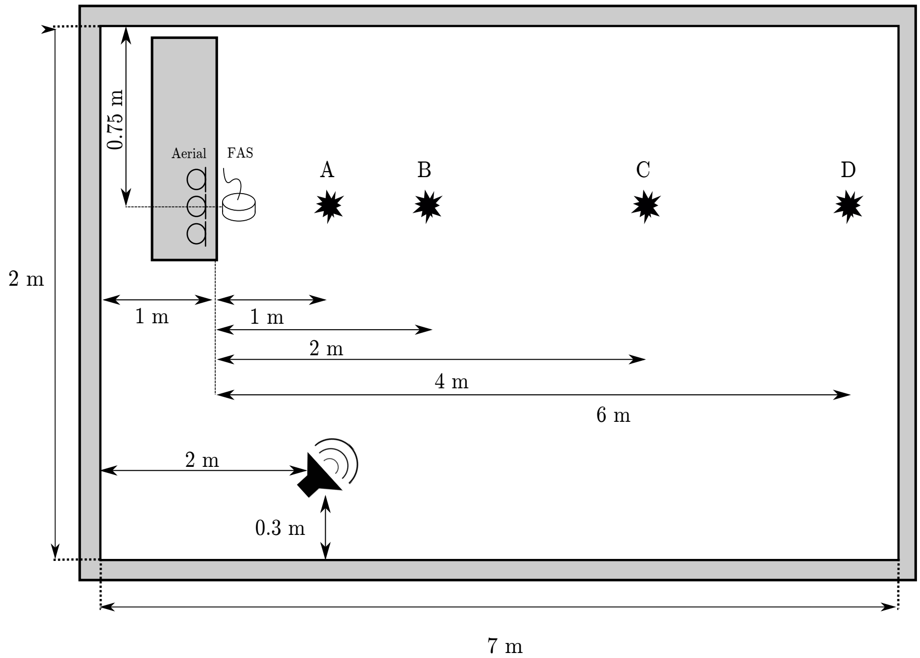
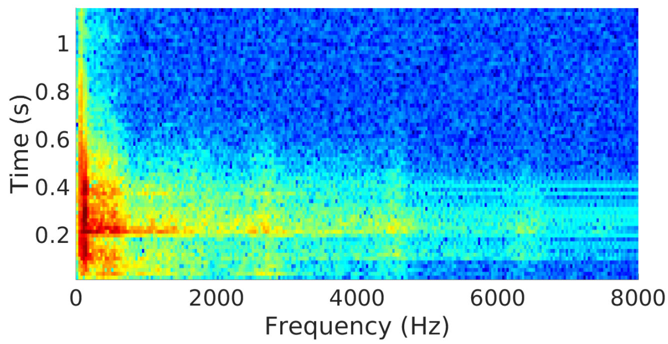


Figure 3.5: The recording room: the letters A, B, C and D indicate the positions of fall events.

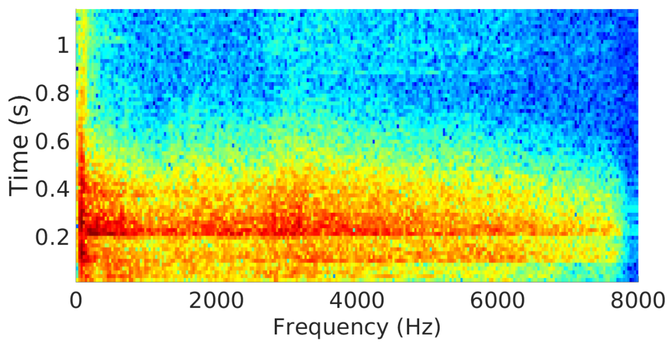
In fact, taking into account the backgrounds tracks as interferences, the floor sensor has a global signal-to-noise ratio (SNR) equal to 20.94 dB and a segmental SNR equal to 7.28 dB. The global SNR of the central aerial microphone is 8.92 dB and the segmental SNR is -1.47 dB. The values of the aerial microphone SNRs are thus considerably lower than the ones of the floor sensor.The global SNR of the floor sensor noisy dataset is 13.66 dB higher than the one of the aerial microphone, highlighting the superior ability of the former to isolate fall signals from external interferences. However, it is worth investigating how the SNR distributes over the frequency range of the acquired signals in order to have a better insight of the physical phenomenon. Figure 3.8 shows the SNR calculated for each mel channel and averaged across the noisy datasets. It can be noticed that the SNR of the floor sensor exceeds the one of the aerial microphone for channels below the fourth. Then, the opposite occurs and the SNR of the aerial microphone assumes greater values. The "valley" in the curves are due to the pitch of the music signal.

QESTO VA NEL LAOVOR ESWA

This analysis suggested the authors that the standard MFCC extraction pipeline can be modified in order to better exploit the properties the floor acoustic sensor. In particular, the analysis led to the following considerations:



(a) Spectrogram of the signal acquired with the floor sensor.



(b) Spectrogram of the signal acquired with the aerial microphone.

Figure 3.6: Frequency content of the same fall event (file "rndy_d2st_bar_0.wav") acquired with the FAS (a) and with the aerial microphone (b).

- The pre-emphasis filter is an inheritance of automatic speech recognition systems, where the input signal is the human voice. Since the effect of the filter is to enhance the high frequency components of the signal and the floor sensor is more sensitive to low frequencies, removing the pre-emphasis could improve the classification performance.
- The majority of the energy of the signals acquired with the floor acoustic sensor is concentrated at frequencies below 1.5 kHz. This suggest that introducing a low-pass filter that removes low SNR frequencies may improve the classification performance.

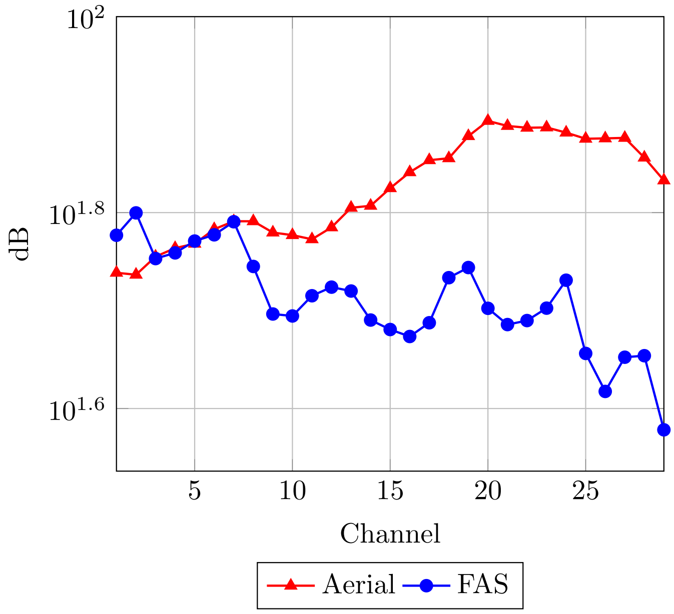


Figure 3.7: Average value of the mel channels.

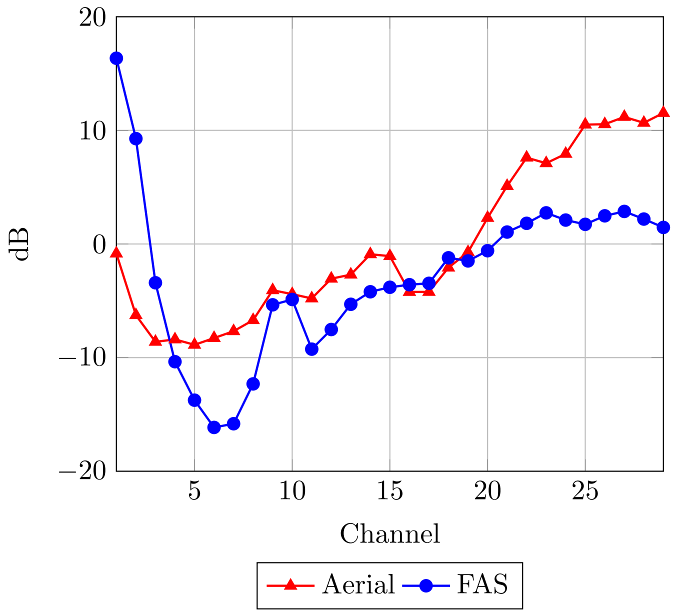


Figure 3.8: Average value of the SNR for each mel channel.

Chapter 4

Supervised Approach

qua vengono presentati sia il metodo multilabel classifier GMM-UBM SVM (ESWN) che il binary GMM-UBM SVM (WIRN2016)

Chapter 5

Unsupervised Approach

Autoencoder wirn 2017 + Qua veine presentato il metodo solo OCSVM

Chapter 6

Semi-Unsupervised Approach

modifica user-aided(CIN) siamese semplice siamese autoencdoer

Chapter 7

Other contributions

List of Publications

[1] R. Bonfigli, A. Felicetti, E. Principi, M. Fagiani, S. Squartini, and F. Piazza, “Denoising autoencoders for non-intrusive load monitoring: Improvements and comparative evaluation,” *Energy and Buildings*, to appear.

[2] R. Bonfigli, E. Principi, M. Fagiani, M. Severini, S. Squartini, and F. Piazza, “Non-intrusive load monitoring by using active and reactive power in additive factorial hidden markov models,” *Applied Energy*, vol. 208, no. Supplement C, pp. 1590 – 1607, 2017.

[3] patent, “Metodo per il monitoraggio non intrusivo del consumo di apparecchiature elettriche collegate ad una linea di alimentazione comune,” Domanda numero: 102017000004554, patent pending.

[4] R. Bonfigli, E. Principi, S. Squartini, M. Fagiani, M. Severini, and F. Piazza, “User-aided Footprint Extraction for Appliance Modelling in Non-Intrusive Load Monitoring,” in *Proc. of the IEEE Symposium Series on Computational Intelligence*, Athens, Greece, Dec. 6-9 2016, pp. 1–8.

[5] R. Bonfigli, M. Severini, S. Squartini, M. Fagiani, and F. Piazza, “Improving the performance of the AFAMAP algorithm for non-intrusive load monitoring,” in *Proc. of the IEEE Congress on Evolutionary Computation (CEC)*, Vancouver, Canada, 2016, pp. 303–310.

[6] M. Fagiani, S. Squartini, R. Bonfigli, M. Severini, and F. Piazza, “Exploiting temporal features and pressure data for automatic leakage detection in smart water grids,” in *2016 IEEE Congress on Evolutionary Computation (CEC)*, July 2016, pp. 295–302.

[7] R. Bonfigli, S. Squartini, M. Fagiani, and F. Piazza, “Unsupervised algorithms for non-intrusive load monitoring: An up-to-date overview,” in *Proc. of IEEE 15th Int. Conf. on Environment and Electrical Engineering (EEEIC)*, June 2015, pp. 1175–1180.

[8] M. Fagiani, S. Squartini, R. Bonfigli, and F. Piazza, “Short-term load forecasting for smart water and gas grids: A comparative evaluation,” in *Environment and Electrical Engineering (EEEIC), 2015 IEEE 15th International Conference on*, June 2015, pp. 1198–1203.

[9] G. Ferroni, R. Bonfigli, E. Principi, S. Squartini, and F. Piazza, "A deep neural network approach for voice activity detection in multi-room domestic scenarios," in *2015 International Joint Conference on Neural Networks (IJCNN)*, July 2015, pp. 1–8.

[10] E. Principi, P. Olivetti, S. Squartini, R. Bonfigli, and F. Piazza, "A floor acoustic sensor for fall classification," in *Audio Engineering Society Convention 138*, May 2015.

[11] E. Principi, S. Squartini, R. Bonfigli, G. Ferroni, and F. Piazza, "An integrated system for voice command recognition and emergency detection based on audio signals," *Expert Systems with Applications*, vol. 42, no. 13, pp. 5668 – 5683, 2015.

[12] G. Ferroni, R. Bonfigli, E. Principi, S. Squartini, and F. Piazza, "Neural networks based methods for voice activity detection in a multi-room domestic environment," in *XIII AI*IA Symposium on Artificial Intelligence*, Dec 2014.

Bibliography

[1] "Eurostat statistic explained: Fertility statistics," March 2016.

[2] G. Carone and D. Costello, "Can europe afford to grow old?," *Finance and Development*, vol. 43, no. 3, pp. 28–31, 2006.

[3] P. Dawadi, D. Cook, and M. Schmitter-Edgecombe, "Automated cognitive health assessment from smart home-based behavior data," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 4, pp. 1188–1194, 2016.

[4] E. Principi, S. Squartini, R. Bonfigli, G. Ferroni, and F. Piazza, "An integrated system for voice command recognition and emergency detection based on audio signals," *Expert Systems with Applications*, vol. 42, no. 13, pp. 5668–5683, 2015.

[5] M. Mubashir, L. Shao, and L. Seed, "A survey on fall detection: Principles and approaches," *Neurocomputing*, vol. 100, pp. 144–152, 2013.

[6] "Department of Health and Human Services: World's older population grows dramatically," `http://www.who.int/en/news-room/fact-sheets/detail/falls`, [Online; accessed 30-Oct-2018].

[7] "World Healt Organization: Falls," `http://www.who.int/en/news-room/fact-sheets/detail/falls`, [Online; accessed 30-Oct-2018].

[8] S. S. Khan and J. Hoey, "Review of fall detection techniques: A data availability perspective," *Medical engineering and physics*, vol. 39, pp. 12–22, 2017.

[9] N. Lapierre, N. Neubauer, A. Miguel-Cruz, A. R. Rincon, L. Liu, and J. Rousseau, "The state of knowledge on technologies and their use for fall detection: A scoping review," *International journal of medical informatics*, 2017.

[10] N. Pannurat, S. Thiemjarus, and E. Nantajeewarawat, "Automatic fall monitoring: a review," *Sensors*, vol. 14, no. 7, pp. 12900–12936, 2014.

[11] T. Xu, Y. Zhou, and J. Zhu, "New advances and challenges of fall detection systems: A survey," *Applied Sciences*, vol. 8, no. 3, pp. 418, 2018.

[12] N. El-Bendary, Q. Tan, F. C. Pivot, and A. Lam, "Fall detection and prevention for the elderly: A review of trends and challenges.," *International Journal on Smart Sensing & Intelligent Systems*, vol. 6, no. 3, 2013.

[13] Q. Wu, Y. D. Zhang, W. Tao, and M. G. Amin, "Radar-based fall detection based on doppler time–frequency signatures for assisted living," *IET Radar, Sonar & Navigation*, vol. 9, no. 2, pp. 164–172, 2015.

[14] T. Liu, H. Yao, R. Ji, Y. Liu, X. Liu, X. Sun, P. Xu, and Z. Zhang, "Vision-based semi-supervised homecare with spatial constraint," in *Pacific-Rim Conference on Multimedia*. Springer, 2008, pp. 416–425.

[15] F. Werner, J. Diermaier, S. Schmid, and P. Panek, "Fall detection with distributed floor-mounted accelerometers: An overview of the development and evaluation of a fall detection system within the project ehome," in *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2011 5th International Conference on*. IEEE, 2011, pp. 354–361.

[16] Y. Zigel, D. Litvak, and I. Gannot, "A method for automatic fall detection of elderly people using floor vibrations and sound—proof of concept on human mimicking doll falls," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 12, pp. 2858–2867, 2009.

[17] Y. Li, K. Ho, and M. Popescu, "A microphone array system for automatic fall detection," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 5, pp. 1291–1301, 2012.

[18] M. Popescu, Y. Li, M. Skubic, and M. Rantz, "An acoustic fall detector system that uses sound height information to reduce the false alarm rate," in *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*. IEEE, 2008, pp. 4628–4631.

[19] M. S. Khan, M. Yu, P. Feng, L. Wang, and J. Chambers, "An unsupervised acoustic fall detection system using source separation for sound interference suppression," *Signal processing*, vol. 110, pp. 199–210, 2015.

[20] X.-X. Zhang, H. Liu, Y. Gao, and D. H. Hu, "Detecting abnormal events via hierarchical dirichlet processes," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, 2009, pp. 278–289.

[21] M. Popescu and A. Mahnot, "Acoustic fall detection using one-class classifiers," in *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*. IEEE, 2009, pp. 3505–3508.

[22] E. E. Stone and M. Skubic, "Fall detection in homes of older adults using the microsoft kinect," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 1, pp. 290–301, 2015.

[23] A. Bourke, J. O'Brien, and G. Lyons, "Evaluation of a threshold-based tri-axial accelerometer fall detection algorithm," *Gait and Posture*, vol. 26, no. 2, pp. 194–199, 2007.

[24] Y. Charlon, N. Fourty, W. Bourennane, and E. Campo, "Design and evaluation of a device worn for fall detection and localization: Application for the continuous monitoring of risks incurred by dependents in an Alzheimer's care unit," *Expert Systems with Applications*, vol. 40, no. 18, pp. 7316–7330, 2013.

[25] A. Özdemir and B. Barshan, "Detecting Falls with Wearable Sensors Using Machine Learning Techniques," *Sensors*, vol. 14, no. 6, pp. 10691–10708, 2014.

[26] M. Lustrek, H. Gjoreski, N. Gonzalez Vega, S. Kozina, B. Cvetkovic, V. Mirchevska, and M. Gams, "Fall detection using location sensors and accelerometers," *Pervasive Computing*, vol. 14, no. 4, pp. 72–79, Oct 2015.

[27] M. Alwan, P. J. Rajendran, S. Kell, D. Mack, S. Dalal, M. Wolfe, and R. Felder, "A smart and passive floor-vibration based fall detector for elderly," in *Proc. of Inf. Commun. Technol.*, 2006, vol. 1, pp. 1003–1007.

[28] A. Yazar, F. Keskin, B. U. Töreyin, and A. E. Çetin, "Fall detection using single-tree complex wavelet transform," *Pattern Recognition Letters*, vol. 34, pp. 1945–1952, 2013.

[29] X. Zhuang, J. Huang, G. Potamianos, and M. Hasegawa-Johnson, "Acoustic fall detection using Gaussian mixture models and GMM supervectors," in *Proc. of ICASSP*, Taipei, Taiwan, Apr. 19-24 2009, pp. 69–72.

[30] L. Liu, M. Popescu, M. Skubic, and M. Rantz, "An automatic fall detection framework using data fusion of Doppler radar and motion sensor network," in *Proc. of the 36th International Conference of the Engineering in Medicine and Biology Society (EMBC)*, 2014, pp. 5940–5943.

[31] C. N. Doukas and I. Maglogiannis, "Emergency fall incidents detection in assisted living environments utilizing motion, sound, and visual perceptual components," *IEEE Trans. Inf. Technol. Biomed.*, vol. 15, no. 2, pp. 277–89, Mar. 2011.

[32] B. Toreyin, A. Soyer, I. Onaran, and E. Cetin, "Falling person detection using multi-sensor signal processing," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, no. 1, pp. 7, 2008.

[33] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[34] C. Bishop, *Pattern Recognition and Machine Learning*, Springer Science+Business Media, LLC, New York, 2006.

[35] B. Schölkopf, R. C. Williamson, A. J. Smola, J. Shawe-Taylor, and J. C. Platt, "Support vector method for novelty detection," in *Advances in neural information processing systems*, 2000, pp. 582–588.

[36] B. Schölkopf, R. C. Williamson, A. J. Smola, J. Shawe-Taylor, and J. C. Platt, "Support vector method for novelty detection," in *Advances in Neural Information Processing Systems*. 2000, vol. 12, pp. 582–588, MIT Press.

[37] S. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge University Press, 2004.

[38] E. C. E. G. S. G. S. Spinsante, "Tst fall detection dataset v1," 2016.

[39] B. Kwolek and M. Kepski, "Human fall detection on embedded platform using depth maps and wireless accelerometer," *Computer methods and programs in biomedicine*, vol. 117, no. 3, pp. 489–501, 2014.

[40] I. Charfi, J. Miteran, J. Dubois, M. Atri, and R. Tourki, "Optimized spatio-temporal descriptors for real-time fall detection: comparison of support vector machine and adaboost-based classification," *Journal of Electronic Imaging*, vol. 22, no. 4, pp. 041106, 2013.

[41] A. T. Özdemir and B. Barshan, "Detecting falls with wearable sensors using machine learning techniques," *Sensors*, vol. 14, no. 6, pp. 10691–10708, 2014.

[42] E. Casilari, J.-A. Santoyo-Ramón, and J.-M. Cano-García, "Analysis of public datasets for wearable fall detection systems," *Sensors*, vol. 17, no. 7, 2017.

[43] M. Vacher, S. Bouakaz, M.-E. B. Chaumon, F. Aman, R. A. Khan, S. Bekkadja, F. Portet, E. Guillou, S. Rossato, and B. Lecouteux, "The cirdo corpus: Comprehensive audio/video database of domestic falls of elderly people," in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, N. C. C. Chair),

K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, and S. Piperidis, Eds., Paris, France, may 2016, European Language Resources Association (ELRA).

[44] P. Olivetti, "Sistema per la rilevazione e prevenzione di caduta anziani, mediante cassa di risonanza a pavimento," Italian Patent 0001416548, July 1, 2015.