Expert Systems With Applications 60 (2016) 51–61

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Acoustic cues from the floor: A new approach for fall classification

Emanuele Principi[a,*], Diego Droghini[a], Stefano Squartini[a], Paolo Olivetti[b], Francesco Piazza[a]

[a] Università Politecnica delle Marche, Via Brecce Bianche, 60131, Ancona, Italy
[b] Scientific Direction, Italian National Institute of Health and Science on Ageing (INRCA), 60124, Ancona, Italy

ARTICLE INFO

Article history:
Received 4 December 2015
Revised 16 March 2016
Accepted 5 April 2016
Available online 27 April 2016

Keywords:
Floor acoustic sensor
Acoustic fall detection
Ambient assisted living
Support vector machine

ABSTRACT

The interest in assistive technologies for supporting people at home is constantly increasing, both in academia and industry. In this context, the authors propose a fall classification system based on an innovative acoustic sensor that operates similarly to stethoscopes and captures the acoustic waves transmitted through the floor. The sensor is designed to minimize the impact of aerial sounds in recordings, thus allowing a more focused acoustic description of fall events. The audio signals acquired by means of the sensor are processed by a fall recognition algorithm based on Mel-Frequency Cepstral Coefficients, Supervectors and Support Vector Machines to discriminate among different types of fall events. The performance of the algorithm has been evaluated against a specific audio corpus comprising falls of a human mimicking doll and of everyday objects. The results showed that the floor sensor significantly improves the performance respect to an aerial microphone: in particular, the F_1-Measure is 6.50% higher in clean conditions and 8.76% higher in mismatched noisy conditions. The proposed approach, thus, has a considerable advantage over aerial solutions since it is able to achieve higher fall classification performance using a simpler algorithmic pipeline and hardware setup.

1. Introduction

In the majority of industrialized countries the ratio between the number of people aged more than 65 years and the number of people aged from 15 to 64 is constantly increasing (Carone & Costello, 2006). As so, it is foreseen that healthcare systems will be put through serious stress in the near future, since this demographic change will result in an increased demand for healthcare services. Governments and public institutions are thus investing in solutions for reducing the burden on healthcare systems. An effective approach is to enable the future homes to take care of its inhabitants by equipping them with technologies able to prevent injuries, detect emergencies, and generally to support them in their everyday life (van den Broek, Cavallo, & Wehrmann, 2010).

In this context, the research community devoted particular attention to solutions for detecting emergencies, in particular persons' falls (Mubashir, Shao, & Seed, 2013). Falls are indeed the primary cause of injury-related death for the elders (Mubashir et al., 2013), and even if they do not result in a physical injury, the more the time a person spends immobile the greater the chance for negative health consequences, both physiological and psychological. The requirements for a prompt detection of persons' falls drove the scientific community to devote several efforts to face this challenging task.

In the literature, fall detection systems employ two types of sensing technologies: environment sensors, such as infrared sensors, pressure, microphones, video sensors, or floor vibration sensors, and wearable sensors, mostly based on accelerometers (Yang & Hsu, 2010). Both types present advantages and disadvantages: environment sensors can suffer from blind spots, require a higher installation cost and are confined to a delimited area. In addition, some sensors, such as cameras, raise privacy concerns. Regarding wearable sensors, they are battery powered, thus while recharging they are not operational, and people may forget wearing the device. Choosing the right technology, thus, is a matter of compromise, depends on the target environment and it is highly subjective: some people may consider intrusive and bothersome wearing a device, thus preferring environmental sensors, while others may prefer the opposite.

In this paper, the presented fall classification system is based on an acoustic environment sensor. In particular, we propose a new sensor that captures the acoustic waves transmitted through the floor as a consequence of a fall. In this sense, the sensor

* Corresponding author.
E-mail addresses: e.principi@univpm.it (E. Principi), d.droghini@pm.univpm.it (D. Droghini), s.squartini@univpm.it (S. Squartini), pa.olo.83@hotmail.it (P. Olivetti), f.piazza@univpm.it (F. Piazza).

http://dx.doi.org/10.1016/j.eswa.2016.04.007
0957-4174/

operates similarly to a stethoscope, with a surface that is in direct contact with the medium transmitting the acoustic waves. The sensor is composed of a microphone embedded in a resonant enclosure whose bottom surface is in direct contact with the floor. In this way, the microphone records the acoustic waves transmitted through the floor and it mainly captures the sound of falling objects, resulting in a minor sensitivity to the disturbances that propagate through the air. Being in direct contact with the floor, the sensor is able to capture subtle signal components, which are absent in the signal transmitted through the air.

The signals acquired with the sensor are classified by means of a fall classification algorithm. The algorithm is based on Mel-Frequency Cepstral Coefficients (MFCCs) (Davis & Mermelstein, 1980) as low-level acoustic features and Gaussian means supervectors (GMS) (Kinnunen & Li, 2010) as features for a Support Vector Machine (SVM) classifier (Cortes & Vapnik, 1995). More in details, a background model is created from a large set of audio signals, and for each audio event class taken into consideration, a set of supervectors is calculated by adapting the background model with the Maximum a Posteriori algorithm and extracting the means of the Gaussians. The supervectors are then employed for training the Support Vector Machine classifier.

The performance of the system has been assessed on a large corpus of fall events created by the authors. The corpus contains human fall events, reproduced by employing “Rescue Randy”[1], a human mimicking doll frequently employed in the fall detection literature (Alwan et al., 2006; Werner, Diermaier, Schmid, & Panek, 2011; Zigel, Litvak, & Gannot, 2009), and general fall events, reproduced by employing six everyday objects (fork, a book, a chair, a basket, a bag, and a ball). Each fall event has been recorded with the floor sensor and with an aerial microphone at four different distances. In order to assess the classification accuracy in noisy conditions, a musical background has been recorded and then digitally added to the fall events.

In a related work by some of the authors (Principi, Olivetti et al., 2015), the system has been described and preliminary experiments have been presented to show the feasibility of the approach. In this paper, the work is extended by conducting a thorough comparison between the performance of an aerial microphone and the one of the proposed sensor, in particular showing the advantages of the proposed solution in realistic noise conditions. An in-depth analysis of the acquired signals has been also conducted, and the MFCC extraction pipeline has been modified accordingly to better exploit the characteristics of the sensor. In addition, in this work a new dataset has been created where human fall events are played by the “Rescue Randy” doll in order to conform to the works on fall detection systems appeared recently on the scientific literature.

The outline of the paper is the following: Section 2 presents an overview of the recent literature on fall detection systems. Section 4 describes the proposed acoustic sensor, analysing the difference between the signals acquired with aerial microphones. Section 5 is devoted to the description of the fall classification algorithm. Section 6 presents the acoustic fall events dataset. Finally, Section 7 presents the experiments conducted to evaluate the performance of the system, and Section 8 concludes the paper and presents future developments.

2. Related contributions

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 (Bourke, O'Brien, & Lyons, 2007) and (Charlon, Fourty, Bourennane, & Campo, 2013) detect a fall by verifying if the acceleration signals exceed a certain threshold. In contrast, in (Özdemir & Barshan, 2014) 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. Lustrek et al. (2015) 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 (Alwan et al., 2006) 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 (Yazar, Keskin, Töreyin, & Çetin, 2013) 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 (Zhuang, Huang, Potamianos, & Hasegawa-Johnson, 2009), 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 (Salman Khan, Yu, Feng, Wang, & Chambers, 2015) 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

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