Signal Processing 110 (2015) 199–210

Contents lists available at ScienceDirect

Signal Processing

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

An unsupervised acoustic fall detection system using source separation for sound interference suppression

Muhammad Salman Khan [a], Miao Yu [a,*], Pengming Feng [a], Liang Wang [b], Jonathon Chambers [a,1]

[a] School of Electronic, Electrical and Systems Engineering, Loughborough University, LE11 3TU, UK
[b] National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences, Beijing, China

ARTICLE INFO

Article history:
Received 1 February 2014
Received in revised form
4 August 2014
Accepted 12 August 2014
Available online 27 August 2014

Keywords:
Health care
Fall detection
Unsupervised classification
Source separation
Mel-frequency cepstral coefficient
One class support vector machine

ABSTRACT

We present a novel unsupervised fall detection system that employs the collected acoustic signals (footstep sound signals) from an elderly person's normal activities to construct a data description model to distinguish falls from non-falls. The measured acoustic signals are initially processed with a source separation (SS) technique to remove the possible interferences from other background sound sources. Mel-frequency cepstral coefficient (MFCC) features are next extracted from the processed signals and used to construct a data description model based on a one class support vector machine (OCSVM) method, which is finally applied to distinguish fall from non-fall sounds. Experiments on a recorded dataset confirm that our proposed fall detection system can achieve better performance, especially with high level of interference from other sound sources, as compared with existing single microphone based methods.

1. Introduction

Currently, there is an increase in aging population across the globe particularly in developed countries. As presented in Carone and Costello [5], the ratio between the number of 65+ people to those between 15 and 64 in the European Union (EU) is projected to double to 54% by 2050 and the topic of home care for elderly people is receiving increasing attention as a consequence. Within the field of health care for elderly people, one important issue is to detect whether an elderly person has fallen or not [9]. As shown in Hsieh et al. [9], falls can cause problems for an elderly person physiologically, such as broken bones, connective and soft tissue damage, even death; although some falls do not result in physical injuries, the elderly people who fall cannot get up without assistance and this period of time spent immobile also affects their health. Due to the serious damage which can be inflicted by falls to elderly people, detection of falls is an important aspect of assisted living. Instead of assigning nurses to monitor whether elderly people fall or not in their homes in a 24/7 manner, an automatic fall detection method is required, which will detect a fall event when it happens so that alarm signals will be sent to certain caregivers (such as hospitals, health centers or relatives) to provide assistance to the elderly person.

Different methods have been proposed for detecting falling activities in recent years. Karantonis et al. [10] proposed a real-time classification system for the types of human movement associated with the data acquired from a single, waist-mounted, triaxial accelerometer unit. In their approach, acceleration signals generated due to gravity and

* Corresponding author.
E-mail addresses: m.s.khan2@lboro.ac.uk (M. Salman Khan), m.yu@lboro.ac.uk (M. Yu), p.feng@lboro.ac.uk (P. Feng), wangliang@nlpr.ia.ac.cn (L. Wang), j.a.chambers@lboro.ac.uk (J. Chambers).

[1] Jonathon chambers is a member of EURASIP.

http://dx.doi.org/10.1016/j.sigpro.2014.08.021

body motion were sampled and processed by certain types of digital filters, and a hierarchical binary structure classifier was then applied on the processed data for classifying different types of movements and detecting falls based on a second-by-second decision process. This system was able to distinguish between periods of activity and rest; recognize the postural orientation of the wearer; and detect events such as walking and falling. According to their experimental results, a fall detection rate of 95.6% was obtained. Instead of fixing the accelerometer at the waist position, Kangas et al. [21] tested the performance of a triaxial accelerometer attached to the subject's body in different positions: head, waist and wrist to detect fall activities. The acceleration information measured by the accelerometer in different positions was compared with an appropriate threshold to determine a fall. The results showed that fall detection using a triaxial accelerometer worn at the waist or head together with a simple threshold-based algorithm is efficient, with a sensitivity of 97–98% and a specificity of 100%. Acceleration sensors can be used with other devices to achieve a comprehensive fall detection system, in Estudillo-Valderrama [6], a low-power waterproof biocompatible accelerometer smart sensor (ACSS) was applied and an additional user interface module was integrated in the second layer (denoted as personal server (PSE) in this paper) to allow the elderly person to access some of the most important data being processed; from the algorithm aspect, an additional time analysis was used by convolving the resulting acceleration data segment with certain defined waveforms, to detect some problematic fall events such as a knee fall. A total of 332 samples of fall and non-fall activities simulated by 31 young and healthy males and females were tested, 100% sensitivity and 95.68% specificity were obtained and a further reduction of false positives can be obtained by manually canceling the fall alarm through the user interface.

Due to advances in computer vision and camera/video and image processing techniques, camera sensor based computer vision methods were also widely applied to detect falls. Some computer vision methods extract video features from the recorded image sequences and the feature values are compared with a certain threshold to determine whether a fall happens or not. Miaou et al. [24,25] proposed a detection system consisting of an omni-dimensional camera and a computer server, which had the advantage of capturing 360° simultaneously in a single shot to remove blind spots. In this approach, a clean background was first obtained. After that, the foreground of interest was obtained by subtracting the background model from the current image and a rectangle enclosing the foreground object was created. The height to width ratio of this rectangle was taken as a feature and compared with a particular threshold to detect falls. The threshold value in this system was customizable depending on the personal physique. The experimental results showed a detection rate of 78% without personal information that increased to 90% with personal information. Rougier et al. [31] proposed a fall detection system based on the motion history image and some changes in the shape of the person. The movement amplitude was measured by the motion history image (MHI) obtained from the frame differencing results and when a large amplitude movement was detected, the shape change feature (such as the changes of the aspect ratio and the orientation angle of the fitted ellipse) was compared with proper thresholds for fall detection. The threshold values were chosen empirically and the experimental results showed a good rate of fall detection with a sensitivity of 88%, and an acceptable rate of false detection with a specificity of 87.5% was obtained, assuming a fixed threshold. Instead of 2D features, some 3D features can be extracted and compared with proper thresholds for fall detection. Auvinet et al. [2] applied calibrated cameras to reconstruct the three-dimensional shape of a person and fall events were detected by analyzing the vertical axis's volume distribution. When the major part of this distribution was abnormally near the floor over a predefined period of time, it is implied that a person had fallen on the floor and an alarm was triggered. The experimental results showed good performance of this system (achieving 99.7% fall detection rate or better with four cameras or more) and a graphic processing unit (GPU) was applied for efficient computation. Considering that sometimes the Euclidean distance between extracted features may not reflect the real semantic similarity between images, Yu et al. [37] propose a novel semantic preserving distance metric learning (SP-DML) algorithm to encode the visual features and semantic contents in a new distance metric construction. The new distance metric could be applied to measure the dissimilarity between two images in a more accurate way by integrating the semantic contents.

Extracted features could also be applied together with a classifier, to classify falls/non-fall activities. Mirmahboub et al. [26] proposed a view-invariant fall detection system by using a single camera. The silhouette area extracted by background subtraction combined with inclination angle was extracted from a video sequence as features. And these were then fed into a support vector machine (SVM) for classifying fall activities and non-fall activities. Different kernels were tested in this work and the experimental results on a public dataset showed that the polynomial kernel of second degree can achieve the best performance with 100% fall detection rate and less than 1% of mistaking non-fall activities as falls. Yu et al. [39] extracted ellipse features and projection histogram features from postures obtained from background subtraction results, and the obtained features were applied to construct a directed acyclic graph support vector machine (DAGSVM) classifier to classify four different types of postures (stand, bend, sit and lie). The classification results, together with the floor region detected during a floor detection phase, were applied to detect falls. The fall detection system was tested on a dataset of 15 people, a high fall detection rate (97.08%) and very low false detection rate (0.8%) were achieved.

We need to notice that it is not always easy to label all the training features for the classifier construction; in order to solve the problem, Yu et al. [38] propose an adaptive hypergraph learning method. The proposed method inherits the advantage of the traditional hypergraph learning method as in Zhou et al. [40], which models the high-order relationship among samples. Besides, compared with the traditional hypergraph learning method, an improved hypergraph construction approach is adopted in