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

PLEASE CITE THE PUBLISHED VERSION

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

PUBLISHER

Crown Copyright © Published by Elsevier B.V.

VERSION

VoR (Version of Record)

REPOSITORY RECORD

Khan, Muhammad Salman, Miao Yu, Pengming Feng, Liang Wang, and Jonathon Chambers. 2014. “An Unsupervised Acoustic Fall Detection System Using Source Separation for Sound Interference Suppression”. Loughborough University. https://hdl.handle.net/2134/16531.

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

Yu et al. [38] by varying the size of the neighborhood for multiple hyperedges construction. And both the labels of unlabeled training samples and the weights of hyperedges could be learned in a simultaneous way to improve the classification performance. Sometimes, multiview features (different types of features such as color, shape, and texture) may be available to be exploited. Liu and Tao [17] and Liu et al. [18] exploit multiple type features for image annotation application by multiview Hessian regularization (mHR)/multiview Hessian discriminative sparse coding (mHDSC) methods respectively, which effectively solve the poor generalization performance problem of the traditional Laplacian regularization (LR) method designed for a particular type of single view features. Yu et al. [36] propose a novel high-order distance-based multiview stochastic learning (HD-MSL) method, which combines multiple types of features into a unified representation and integrates the labeling information based on a probabilistic framework. The proposed HD-MSL method can automatically learn a combination coefficient of different types of features, which can exploit complementary information and by the aid of the alternative optimization, the classification scores are obtained simultaneously. The experimental results presented the effectiveness of the HD-MSL method compared with other state-of-the-art ones with much higher classification accuracy. In summary, the methods in Yu et al. [38], Liu and Tao [17], Liu et al. [18] and Yu et al. [36] have potentials to be applied for building up a more efficient classifier for classifying falls/non-fall activities from multiview, partly labeled features.

For the accelerometer sensor based method, there is a need for the elderly person to wear the accelerometer sensor so that it is both obtrusive and inconvenient. For the computer vision based method, there is a privacy issue due to the fact that the daily life of an elderly person is recorded by a camera, but this is unlikely to deter an elderly person using the system if it preserves their independence. Besides, change in a room environment, such as the movement of furniture and illumination change, will affect the performance of the computer vision based method. In order to overcome these limitations, the acoustic sensors (microphones) could be applied and we therefore propose a novel fall detection system based on acoustic sensors in this work. Compared with other related acoustic sensor(s) based methods, our proposed technique is more robust to background acoustic noises because a novel SS technique removes the possible interferences from other sound sources by using only two microphones. Considering the difficulty to obtain the actual falling sounds to build up the supervised two class classifier, in our proposed method, only non-fall sounds from normal activities are applied to build a data description model, which can effectively distinguish falling or non-fall sounds. The organization of this paper is as follows: Section 2 gives a brief overview of the proposed acoustic sensor based fall detection system and other related work; the implementation details of this proposed fall detection system are given in Section 3. Section 4 describes the evaluation of the proposed fall detection system and the final discussions and conclusions are provided in Sections 5 and 6.

2. Overview of the proposed fall detection method and other related works

In order to overcome the limitations of the accelerometer based methods and computer vision based methods, some researchers use acoustic sensors (microphones) to detect falls based on the acquired audio signal. In Zigel et al. [41], a fall detection system based on a floor vibration sensor and a microphone was proposed; the vibration and sound signals were obtained from the hardware equipment and temporal and spectral features were extracted from the resulting signals. Bayes' classifier was then applied to classify fall and non-fall activities based on the extracted features. In their work, a doll which mimicked a human was used to simulate falls and their system detected such falls with a fall detection rate of 97.5% and a false detection rate of 1.4%. In Li et al. [16] an acoustic fall detection system was developed, which automatically detected a fall and reported it to the caregiver. The study used an 8-microphone circular array which provided a better 3-D estimation of the sound location by using the steered response power with the phase transform (SRP-PHAT) algorithm Mungamuru and Aarabi [27], and the sound signal was then enhanced by a beamforming technique with the aid of the resulting location information. Mel frequency cepstral coefficient (MFCC) features were extracted from the enhanced sound signal and the *k*th nearest neighbor method was applied to discriminate a fall from a non-fall activity. A pilot experiment on a dataset containing 30 fall activities and 120 non-fall activities was performed, all the falls were detected and only six non-fall activities were taken as fall activities. An improvement of Li et al. [16] was proposed in Li et al. [15] by introducing height information for the sound source estimated by an eight-microphone array; if a sound source's height was larger than a particular threshold, then it was unlikely to be a fall. In this way, the false alarms due to background noise were reduced to a large extent. A larger dataset which contained 120 simulated fall sounds and 120 simulated non-fall sounds generated by three stunt actors was used for evaluation. A good performance was obtained with 100% fall detection rate and 3% false detection rate. In order to solve the problem that it is not easy to obtain realistic fall sound for training, a one class classifier technique was proposed in Popescu and Mahnot [29] using only the non-fall sound recorded from a single microphone for one class classifier construction. The sound signals were initially pre-processed by a Wiener filter for noise removal and the extracted MFCC features were then used to build the corresponding classifiers. In this work, three types of one class classifier: nearest neighbor classifier, the one class support vector machine classifier and mixture of Gaussian classifier were tested. From the preliminary results, it was found that the fall detection results achieved by the three one class classifiers were comparable with those by using the popular two-class support vector machine in terms of the ROC curve analysis.

A new acoustic sensors based fall detection method is proposed in our work, which applies a novel SS technique to remove background noises; besides, the one class support vector machine (OCSVM) technique is applied and only the MFCC features from non-fall sounds are applied to construct the OCSVM data description model to distinguish falling/non-fall sounds. The block diagram of the proposed two-microphone based fall detection system is shown in Fig. 1; initially, an elderly person undertakes normal activities in a room environment and the generated acoustic signals (footstep sounds) are collected. The system is targeted at the situation where a single person lives alone and interference by pets is beyond the scope of this study. For each sequence of the collected acoustic signal, the method in Loesch and Yang [19] is applied to estimate the number of sound sources. If the estimated number of sound sources is larger than one, which implies the acoustic signal is mixed with some interference (such as a mixture of footstep sounds and TV sounds), the SS technique proposed in Khan et al. [12] is applied to separate the mixed acoustic signals to obtain the individual signal coming from each source. Moreover, the position of each sound source is also estimated, which can be taken as a cue to determine the interferences (such as the most common TV interference in a real home environment as mentioned in [41]) and can be reinforced by the fact that the prior position knowledge of an interference source (usually the TV position is fixed and can be obtained in a real home environment). The MFCC features Ganchev et al. [7] corresponding to the non-interference acoustic signals are next extracted and applied to construct a data description model based on a one class support vector machine (OCSVM) technique. This constructed OCSVM normal model is finally applied to distinguish normal sounds (footstep sounds generated by normal activities) from abnormal sounds (falling sounds).

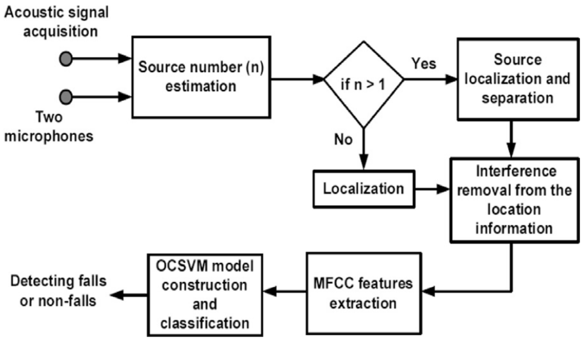
Acoustic signal acquisition
Two microphones
Source number (n) estimation
if n > 1
Yes
No
Source localization and separation
Localization
Interference removal from the location information
MFCC features extraction
OCSVM model construction and classification
Detecting falls or non-falls

Fig. 1. The block diagram of the proposed fall detection system.

3. Methods

Before introducing our proposed method, we clarify important notations in Table 1. These notations are used when we present the main steps of the proposed method, which include the interference suppression, MFCC feature extraction and OCSVM modelling.

3.1. Interference suppression

Initially, we use the method in Loesch and Yang [19] to estimate the number of sound sources from a recorded acoustic signal. The method operates in the time–frequency (TF) domain by first identifying reliable TF points, estimating the direction-of-arrival and then clustering them using nearest-neighbor classification. If the number of the estimated sources is one, then we determine whether the acoustic signal is interference or not by the source position estimated by the source localization scheme [23]. The acoustic signal is regarded as an interference if the estimated source location is from a known interference source position (such as TV position) and is thereby discarded.

If the number of estimated sources is more than one, a two-stage source separation method utilizing only two microphones is performed. Since, in general, most realistic enclosures are highly reverberant, i.e. reverberation time (RT60) is over 400–500 milliseconds, therefore, in the first stage we dereverberate the observed two-channel acoustic mixture. The dereverberation scheme is based on spectral subtraction [12]. In the second stage we employ an efficient model-based source separation technique which is motivated by aspects of the human auditory system by

Table 1
List of important notations.

Main steps	Notation	Description
Interference suppression and MFCC feature extraction	$X_{rev}(f,t)$	Observed reverberant signal
	$X_{rev_{late}}(f,t)$	Late reverberant component
	$X_{cln}(f,t)$	Clean signal
	$G(f,t)$	Gain function applied to yield the clean signal
	$SIR_{post}(f,t)$	*Aposteriori* signal-to-noise ratio
	$\mathcal{N}(\cdot \vert \mu(f),\eta^2(f))$	Normal distribution with mean $\mu(f)$ and variance $\eta^2(f)$
	$\mathbf{y}(f,t)$	Mixture at frequency f and time instance t
	$\mathbf{h}_d(f)$	Mixing vector
	$\mathbf{d}_s(f)$	Direction vector
	$\varsigma_s^2(f)$	Model variance
	$H_i(k)$	Frequency response of the ith triangular filter at the frequency k
OCSVM modelling	$f(\mathbf{x})$	Hyperplane function estimated by OCSVM
	$\mathbf{w},\rho$	Parameters determining the hyperplane
	$\Phi(\mathbf{x})$	Mapped vector $\mathbf{x}$ in a feature space
	α	Dual variables
	$k(\cdot,\cdot)$	Kernel function

combining the models of interaural level difference (ILD), interaural phase difference (IPD) and the model of mixing vectors [11]. This probabilistic modeling, which is performed in the time–frequency (TF) domain, yields TF soft masks for each source in the mixture that can be used for their reconstruction.

3.1.1. Dereverberation

A sound source in general reaches the microphones by following a direct path and multiple reflective paths. In a realistic room these reflections could be on the order of thousands of samples in duration with typical sampling rates, making the observed acoustic signal highly reverberant. The late part of such reverberation is said to have particularly detrimental effect. Thus, in the first stage, we dereverberate the measured two-channel acoustic mixtures. We achieve this by using a spectral subtraction based binaural dereverberation method [12].

If, in the TF domain, $X_{rev}(f,t)$ is the observed reverberant signal and $X_{rev_{late}}(f,t)$ is the late reverberant component, then the clean signal $X_{cln}(f,t)$ at frequency index f and time frame t for spectral subtraction based dereverberation methods can be written as $X_{cln}(f,t) = X_{rev}(f,t) - X_{rev_{late}}(f,t)$. The process can also be written as

$$X_{cln}(f,t) = G(f,t)X_{rev}(f,t) \tag{1}$$

where $G(f,t)$ is a gain function applied to the observed reverberant signal to yield the dereverberated clean signal. The gain is estimated as [14]

$$G(f,t) = 1 - \frac{1}{\sqrt{SIR_{post}(f,t)+1}} \tag{2}$$

where $SIR_{post}(f,t) = |X_{rev}(f,t)|^2 / \sigma^2_{X_{rev_{late}}}(f,t)$ is the *a posteriori* signal-to-noise ratio (SNR) and $|.|$ denotes the magnitude operation. Here $\sigma^2_{X_{rev_{late}}}(f,t)$ is the variance of the late reverberant speech component. The left and right channel reverberant signals are independently processed to obtain the gain function as explained above. The two gains are combined [12] to form a single gain that is applied to the mixtures for dereverberation. The dereverberated mixtures are supplied to the second stage for separation.

3.1.2. Source localization and separation

The ratio of the left and right dereverberated signals in the TF domain gives the observed ILD and the IPD at frequency f and time t. The phase residual, the difference between the observed IPD and the predicted IPD (by a delay of τ samples), is modeled with a normal distribution with frequency-dependent mean $\xi(f)$ and variance $\sigma^2(f)$ [23], $p(\phi(f,t)|\tau(f),\sigma(f)) = \mathcal{N}(\hat{\phi}(f,t;\tau)|\xi(f),\sigma^2(f))$. The ILD is also modeled with a normal distribution with mean $\mu(f)$ and variance $\eta^2(f)$ [23], $p(\alpha(f,t)|\mu(f),\eta^2(f)) = \mathcal{N}(\alpha(f,t)|\mu(f),\eta^2(f))$. The left and right channels, in the TF domain, are concatenated to form a mixture $\mathbf{y}(f,t)$. Assuming the signals are sparse in the TF domain [35] and only one source is dominant at each TF point, $\mathbf{y}(f,t)$ at each time t and frequency f can be approximated as [32] $\mathbf{y}(f,t) \approx \mathbf{h}_d(f)s_d(f,t)$, where $\mathbf{h}_d(f) = [h_{ld}(f), h_{rd}(f)]^T$ is the mixing vector from the dominant source $s_d(f,t)$ to the left and the right sensor at that TF point. The mixing vectors are modeled for each source with a Gaussian model as [32],

$$p\left(\mathbf{y}(f,t)\middle|\mathbf{d}_s(f),\varsigma_s^2(f)\right) = \frac{1}{\pi\varsigma_s^2(f)}\exp\left(-\frac{\|\mathbf{y}(f,t)-(\mathbf{d}_s^H(f)\mathbf{y}(f,t)).\mathbf{d}_s(f)\|^2}{\varsigma_s^2(f)}\right) \tag{3}$$

where $\mathbf{d}_s(f)$ is the direction vector, $\varsigma_s^2(f)$ is the variance of the model, $(\cdot)^H$ is the Hermitian transpose, and $\|\cdot\|$ indicates the Euclidean norm operator.

The ILD, IPD and mixing vector models are combined and the model parameters are estimated in the maximum likelihood sense using iterative expectation–maximization (EM) [11]. TF masks are generated after a fixed number of iterations and then used to reconstruct the acoustic signals from different sources. Together with each reconstructed acoustic signal, its corresponding source position is also estimated by the method in Mandel [23] as in the single source case, in order to determine whether the reconstructed acoustic signal is an interference or not; only the non-interference acoustic signals are retained for further processing.

The time required to run the source separation algorithm, in order to suppress potential interferences, is linear in the number of points within the time–frequency representation, the number of sound sources, the number of discrete values of the delay τ that are used, and the number of EM iterations.

3.2. MFCC feature extraction

Features are next extracted from the non-interference acoustic signals after the interference suppression procedure has been completed. The most commonly used acoustic features for speech/audio recognition are Mel-scale frequency cepstral coefficients (MFCCs). As mentioned in Li et al. [15], MFCCs take into consideration human perception sensitivity with respect to frequencies. In fall detection, we use MFCCs as features to distinguish fall and non-fall sounds, as humans can distinguish such sounds. For an acoustic signal, the procedure of extracting the MFCCs is divided into the following steps as presented in Ganchev et al. [7]:

1. *Segmentation and Hamming windowing*: For an input acoustic signal, it is segmented into frames with approximate 50% overlap. Each frame is multiplied by a Hamming window in order to minimize the boundary effect due to segmentation.

2. For each frame, the discrete Fourier transform (DFT) is applied to convert the time-domain points into the frequency domain, and the magnitude values of the DFT for frames are calculated.

3. Converting the DFT data into filter bank outputs, as presented in Ganchev et al. [7], the filter bank contains 40 equal area triangular filters, which cover the frequency range 133,6854 Hz. The center frequencies of the first 13 are linearly spaced in the range [200,1000] Hz with a step of 66.67 Hz and the next 27 are logarithmically spaced in the range [1071, 6400] Hz with a step $logStep = 1.07$. Each

one of these equal area triangular filters is defined as

$$H_i(k) = \begin{cases} 0 & \text{for } k < f_{i-1} \\ \dfrac{2(k-f_{i-1})}{(f_i - f_{i-1})(f_{i+1} - f_{i-1})} & \text{for } f_{i-1} \le k \le f_i \\ \dfrac{2(f_{i+1}-k)}{(f_{i+1} - f_i)(f_{i+1} - f_{i-1})} & \text{for } f_i \le k \le f_{i+1} \\ 0 & \text{for } k > f_{i+1} \end{cases} \tag{4}$$

where i stands for the ith triangle filter, which is determined by frequencies f_{i-1}, f_i and f_{i+1} (the center frequencies for the $(i-1)$ th, ith and $(i+1)$ th filters). And k corresponds to the kth coefficient of the DFT coefficient.

4. We calculate the log values (with the base 10) of the filter bank outputs and the DCT transform [15] is applied on the outputs, a certain number of DCT coefficients form the final MFCC features.

Most of the computational costs of the MFCC features extraction come from steps (2) and (4), which involve transforming the discrete time domain audio signal into the frequency domain and the discrete cosine transformation of filter bank outputs. Algorithms which accelerate the DFT and DCT, such as the fast Fourier transform (FFT) and fast DCT transform as proposed in Proakis and Manolakis [30] could be applied to reduce the computational costs. By the aid of the fast Fourier transform and fast DCT transform, the arithmetical operations for a N points sequence could be reduced from $\mathcal{O}(N^2)$ to $\mathcal{O}(N \cdot \log_2 N)$, provided N is a power of 2.

From the above four steps, we extract the MFCC features for an input acoustic signal and the MFCC features are then used to construct the OCSVM model for describing non-fall sounds from normal activities, as presented in the next section.

3.3. One class support vector machine

To model the extracted MFCCs, the OCSVM is applied, which is an elegant data fitting method described in Scholkopf et al. [33]. The basic idea behind the OCSVM model is that given a data set $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N]$ drawn from an underlying probability distribution $\mathbf{P}$, the corresponding supporting region can be obtained by estimating a function f. If a sample is obtained from the supporting region, both the distribution probability value and the function f value are large; otherwise, small values of distribution probability and f are obtained.

Compared with the single Gaussian model or mixture of Gaussian model, OCSVM is more flexible because there is no assumption in OCSVM that the data to be fitted should follow particular types of distributions (single Gaussian or a mixture of Gaussians), which are insufficient for high-dimensional data (such as the MFCC features used in this work) description due to the curse-of-dimensionality.

As proposed in Scholkopf et al. [33], the function f is in a linear form as

$$f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} - \rho \tag{5}$$

for the data sample $\mathbf{x}$ (here $\cdot$ represents the dot-product). And in most cases, the data $\mathbf{x}$ is mapped into a feature space with $\mathbf{x} \to \Phi(\mathbf{x})$ in order to obtain a better non-linear result for data fitting and Eq. (5) becomes

$$f(\mathbf{x}) = \mathbf{w} \cdot \Phi(\mathbf{x}) - \rho \tag{6}$$

In order to obtain the parameters $\mathbf{w}$ and ρ, the following quadratic problem needs to be solved based on the dataset $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N]$:

$$\min_{\mathbf{w},\xi,\rho} \ \frac{1}{2}\|\mathbf{w}\|^2 + \frac{1}{\nu N}\sum_i \xi_i - \rho$$
$$\text{subject to} \quad (w \cdot \Phi(\mathbf{x}_i)) \ge \rho - \xi_i, \ \ \xi_i \ge 0 \tag{7}$$

where $\nu \in (0,1]$ and the nonzero slack variables $\xi = [\xi_1, \ldots, \xi_N]$ are introduced to allow for the possibility of outliers (the data points which are not drawn from the supporting region).

Using multipliers $\alpha_i, \beta_i \ge 0$, where $i = 1, \ldots, N$, a Lagrangian function is introduced as

$$L(\mathbf{w}, \xi, \rho, \alpha, \beta) = \frac{1}{2}\|\mathbf{w}\|^2 + \frac{1}{\nu N}\sum_i \xi_i - \rho - \sum_i \alpha_i((\mathbf{w} \cdot \Phi(\mathbf{x})) - \rho + \xi_i) - \sum_i \beta_i \xi_i \tag{8}$$

where $\alpha = [\alpha_1, \ldots, \alpha_N]$ and $\beta = [\beta_1, \ldots, \beta_N]$; the derivatives of the above Lagrangian function with respect to $\mathbf{w}$, ξ and ρ are set to zeros, which yields

$$\mathbf{w} = \sum_i \alpha_i \Phi(\mathbf{x}_i)$$
$$\alpha_i = \frac{1}{\nu N} - \beta_i \le \frac{1}{\nu N}$$
$$\sum_i \alpha_i = 1 \tag{9}$$

The results of (9) are substituted into (8) while considering the constraints of α, as mentioned in Boyd and Vandenberghe [4], and a dual form of problem (7) is obtained as:

$$\min_{\alpha} \ \frac{1}{2}\sum_{ij} \alpha_i \alpha_j \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)$$
$$\text{subject to} \quad 0 \le \alpha_i \le \frac{1}{\nu N}, \ \ i = 1, \ldots, N, \ \ \sum_i \alpha_i = 1 \tag{10}$$

The problem (10) is a convex problem and can be solved by the standard algorithm for solving the convex problem as mentioned in Boyd and Vandenberghe [4]. Instead of representing $\Phi(\mathbf{x})$ explicitly, the kernel technique [3] is applied and a kernel function $k(\mathbf{x}, \mathbf{y}) = \Phi(\mathbf{x}) \cdot \Phi(\mathbf{y})$ is used to represent the dot product of samples in the feature space and (10) is then rewritten as

$$\min_{\alpha} \ \frac{1}{2}\sum_{ij} \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j)$$
$$\text{subject to} \quad 0 \le \alpha_i \le \frac{1}{\nu N}, \ \ \sum_i \alpha_i = 1 \tag{11}$$

And in this work, the popular Gaussian kernel $k(\mathbf{x}, \mathbf{y}) = e^{(-\|\mathbf{x}-\mathbf{y}\|^2)/2\sigma^2}$ is applied, where σ is the Gaussian kernel parameter.

From Eq. (9), the solution of α in (11) (denoted as α^*) is related to the solution of $\mathbf{w}$ in (7) (denoted as $\mathbf{w}^*$) with

$$\mathbf{w}^* = \sum_i \alpha_i^* \Phi(\mathbf{x}_i) \tag{12}$$

The solution of parameter ρ in (7) (denoted as ρ^*) can be found from the Karush–Kuhn–Tucker (KKT) conditions as described in Boyd and Vandenberghe [4], from which the following equations hold:

$$\alpha_i^*((\mathbf{w}^* \cdot \Phi(\mathbf{x})) - \rho^* + \xi_i^*) = 0$$
$$\beta_i^* \xi_i^* = 0 \tag{13}$$

where ξ_i^* and β_i^* denote the ith solutions of ξ and β for minimizing (7).

It can be observed from (13) that for a particular index i, if α_i^* and β_i^* are non-zero, the corresponding data sample $\mathbf{x}_i$ satisfies

$$\rho^* = (\mathbf{w}^* \cdot \Phi(\mathbf{x}_i)) \tag{14}$$

and from Eq. (9), $\mathbf{w}^*$ is replaced with $\sum_i \alpha_i^* \Phi(\mathbf{x}_i)$, then

$$\rho^* = \sum_i \alpha_i^* (\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_i))$$
$$= \sum_j \alpha_j^* k(\mathbf{x}_j, \mathbf{x}_i) \tag{15}$$

Finally, after α^* and ρ^* are obtained, the decision function (6) is determined as

$$f(\mathbf{x}) = \sum_j \alpha_j^* k(\mathbf{x}_j, \mathbf{x}) - \rho^* \tag{16}$$

which is then used to model the dataset $\mathbf{X}$ (MFCC features in this work); if the value of $f(\mathbf{x})$ for a data sample $\mathbf{x}$ is large, then the sample $\mathbf{x}$ is likely to come from the supporting region where most samples of $\mathbf{X}$ reside.

We need to remark that with the aid of the OCSVM algorithm, only the supporting vector (the sample $\mathbf{x}_i$ whose corresponding coefficient $\alpha_i^* > 0$) is kept for constructing the decision function in (6). In this way, the decision function is not represented by all the training samples as in the traditional kernel based method [3] and the computational cost for deciding whether a data sample is normal/abnormal could then be reduced. After training the OCSVM classifier, the computational cost for testing a date sample by the trained OCSVM classifier is $O(N)$, where N is the number of support vectors.

4. Experimental results

4.1. Experimental settings

The system evaluation experiments were performed in a lab environment with the dimension 5.6 m × 4.4 m × 3.5 m as shown in Fig. 2. Two RV 6 microphones were used to record the audio signals and we used a Firepod microphone preamplifier to convert the analog signals to digital ones with a sampling rate of 16 kHz. Matlab R2010b was then used to process the converted digital signals for interference suppression, extraction of MFCCs and classification operations. A television was introduced as the acoustic interference source, which was used to test the performance of the proposed system under the influence of acoustic interferences as in the real-home environment. When a person simulates fall activities, a mattress was used to prevent the person from being injured.

4.2. Comparison with other source separation methods for interferences suppression

To test the efficacy of the proposed two-channel source separation approach, we first compared our algorithm with three other state-of-the-art source separation methods in a simulated environment, as shown in Fig. 3. Speech data, assumed to be the television interference, was mixed with human walking sound at different levels of reverberation. The level of reverberation was varied using the well-known source image method [1]. The synthetic room impulse responses (RIRs) generated using the image method were convolved with the sound sources to model the reverberant sources, which were then mixed to form the reverberant mixtures.

The objective of this experiment was to separate the walking sound from the acoustic mixture and to effectively

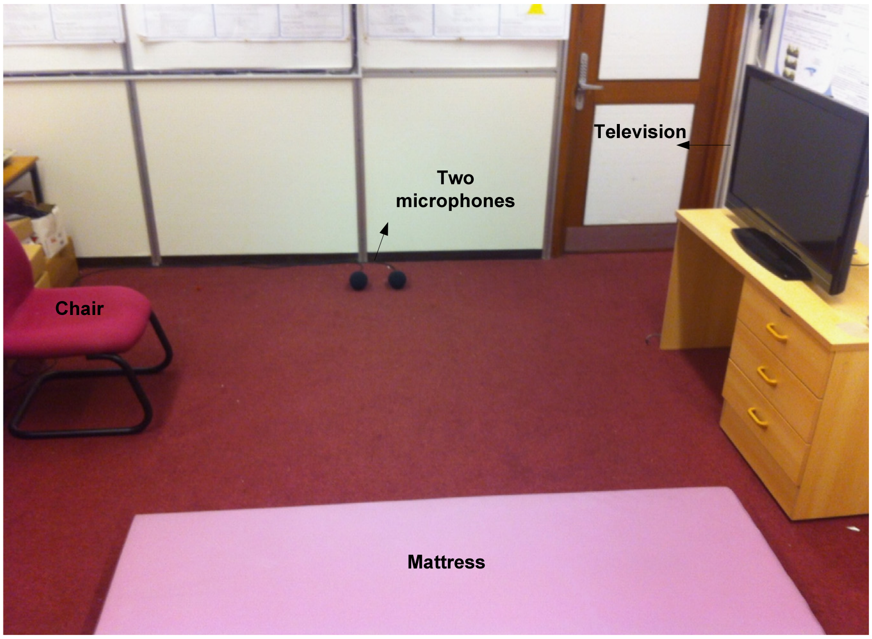

Fig. 2. The experimental room environment with two microphones, and a television (TV) to simulate interference.

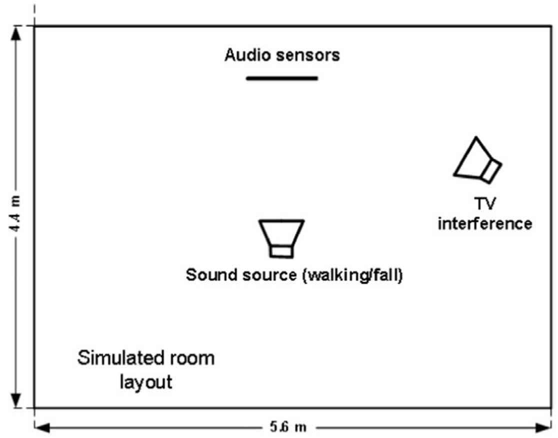

Fig. 3. The simulated room setting with approximate sound source positions. Human walking sound was mixed with the television (TV) interference at different levels of reverberation. The assumed spacing between the audio sensors (microphones) for the proposed method was 0.17 m and the other techniques was 0.04 m. The mixtures were then separated using the different methods and the performance recorded.

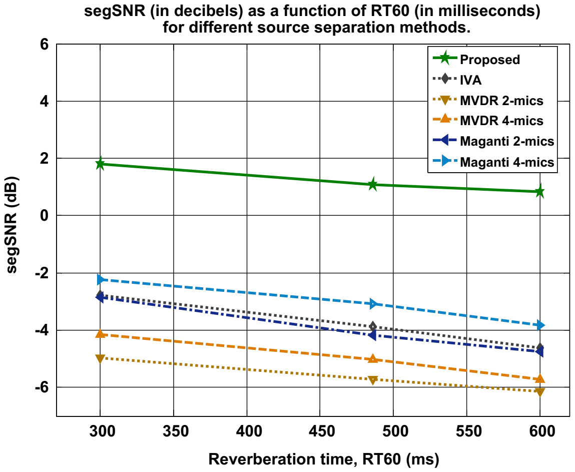

Fig. 4. Comparison of segSNR (in decibels) performance as a function of RT60 using the proposed and IVA algorithms with two microphones and the Naqvi and Maganti methods employing two and four microphones for mixtures of two sources.

measure and compare the separation performance of the proposed source separation scheme with other methods. We chose to use synthetic RIRs and available sound sources in order to be able to quantitatively measure the separation performance using state-of-the-art objective metrics. The frame-based segmental signal-to-distortion ratio (segSNR) [20] was used to evaluate the separation performance of the different methods. Our source separation method was compared with the method in Kim et al. [13] based on independent vector analysis, the method in Naqvi et al. [28] that is based on minimum variance distortionless response (MVDR) beamforming [8] and the technique in Maganti et al. [22] which is also based on array beamforming. It is highlighted that the beamforming based methods in Naqvi et al. [28] and Maganti et al. [22] require the knowledge of the sound source locations, which we assume are known.

The different source separation methods were evaluated at three different reverberation times (RT60) i.e. 300 ms, 485 ms and 600 ms, achieved by varying the reflectivity of the walls. These RT60s were chosen since a typical real home room has an RT60 in this range. The results in Fig. 4 show improved performance by the proposed method over all the other methods at all RT60s. The proposed method with dereverberation based pre-processing effectively dereverberates the mixture which is later separated by the combined models of the interaural level difference (ILD), interaural phase difference (IPD) and mixing vectors. The technique in Maganti utilizing four microphones performs second best. Overall, Maganti's scheme, both with two and four microphones, does better than the simple MVDR beamformer due to the fact that it employs additional post-processing on the beamformers output to further suppress the interference. In summary, the proposed source separation method using only two microphones, with the pre-processing to tackle the high levels of reverberation present in realistic environments, performs better than the beamforming-based methods, even with the *a priori* knowledge of the sound source locations, using two and four microphones in similar conditions.

4.3. Fall detection performance evaluation

To construct the dataset of the normal sound samples, we interviewed a healthy 75 years old person and the frequency of representative activities during 1 week as summarized in Table 2. In the experiments, one volunteer is asked to simulate these activities, and for each activity a sound sequence of 10 s (s) is recorded (which is sufficient to cover the period of a particular activity).

For each acoustic sequence, it was segmented into 1 s sound blocks with 50% overlap, and the MFCC features were then extracted from them to construct the normal OCSVM model, which is used to classify fall and non-fall sounds. For the MFCC features extraction, each 1 s sound block was divided into frames, the length of each frame is set to be 256 points, the overlapping rate between frames was set to be 37.5% and the initial 13 coefficients of the DCT of the outputs of 40 area triangular filters corresponding to each frame were chosen as MFCC features. With a sampling rate of 16 kHz, for one particular sound sample, an MFCC feature vector with a dimension of 1274 was extracted.

Another dataset was recorded which contains the sound sequences of 64 falls (including frontal fall, side fall and backward fall) and 64 non-falls (including walking, standing, bending, lying and sitting) as the test dataset, each sound sequence also lasts for 10 s. For each testing sequence, it was also segmented into 1 s testing sound blocks (with 50% overlap), the extracted MFCC features from the testing sound samples were fed into the OCSVM model for classifying.

To test the influence of the TV interferences, we introduced 50% of such interference (50% interference implies the interference signal variance is one half of that of the target signal variance) on both training and testing datasets

Table 2
Summary of the frequency of representative activities of an elderly person during 1 week.

Activity	Description	Frequency
Walking	The elderly person walks to move between different places of the room or do some cleaning activities	16
Standing	The elderly person stands almost still to watch television	6
Sitting	The elderly person sits to have a rest (either watching TV or eating fruits)	8
Lying	The elderly person lies on the sofa for a nap or watching TV	8

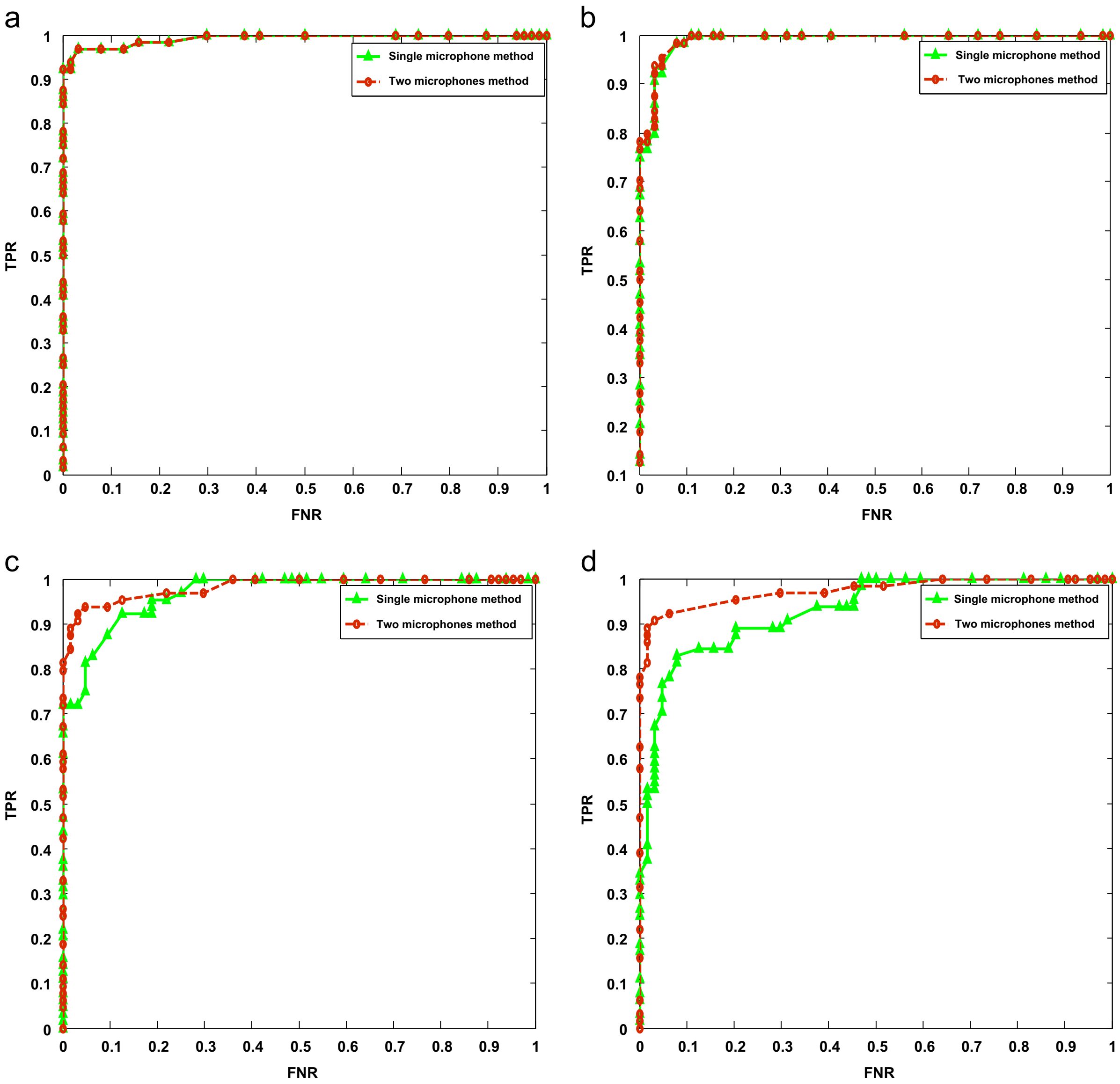

Fig. 5. Comparisons of the ROC curves for fall detection by using single microphone and two microphones under different interference levels: (a) no interference, (b) 25% interference level, (c) 50% interference level and (d) 75% interference level.

(to simulate the situation that a person does normal activities or falls while there is some other sound interference, such as the TV is on). In our work, three different interference levels were tested, they were 25%, 50% and 75% of the maximum volume of the laptop.

Receiver operating characteristic (ROC) analysis [34] was applied for evaluation purpose. Different thresholds were chosen for the OCSVM model and the true positive rate (TPR, which represents the percentage of falls which are correctly detected) and false negative rate (FNR, which

Table 3
The performance of the single microphone method [41,29] under different interference levels.

Evaluation criteria	No interference	25% interference level	50% interference level	75% interference level
AUC value	0.9928	0.9902	0.9684	0.9293
optimal TPR	0.9688	0.9844	0.9219	0.8281
optimal FNR	0.0313	0.0781	0.1250	0.0781
Geometric mean	0.9687	0.9526	0.8981	0.8737

Table 4
The performance of the proposed method under different interference levels.

Evaluation criteria	No interference	25% interference level	50% interference level	75% interference level
AUC value	0.9928	0.9912	0.9829	0.9738
optimal TPR	0.9688	0.9531	0.9375	0.9063
optimal FNR	0.0313	0.0469	0.0419	0.0313
Geometric mean	0.9687	0.9531	0.9477	0.9370

represents the percentage of non-falls which are wrongly detected as falls) were calculated for these thresholds, and the results are plotted as a ROC curve. Ideally, for a perfect fall detection system, TPR should be 1 and FNR should be 0.

As proposed in Tax [34], two criteria can be obtained from the ROC curve for performance evaluation, they are:

1. AUC value – which denotes the area under the ROC curve, a larger AUC value means a better performance of the corresponding model used for detecting falls.

2. Optimal TPR and FNR pair under a particular threshold, which maximizes the geometric mean, $\sqrt{TPR * (1-FPR)}$, whose range is [0, 1], for a perfect system with unity TPR and FNR zero the corresponding value of geometric mean is unity.

Fig. 5 shows the ROC curves of our proposed fall detection system for four scenarios (without interference, and 25%, 50% and 75% of maximum volume interference), the corresponding AUC, optimal TPR and FNR pair and optimal geometric mean values are summarized in Table 4. For comparison purpose, we also give the performance of the fall detection system by using only one microphone as in Zigel et al. [41] and Popescu and Mahnot [29] without interference suppression and the results are summarized in Table 3. From these two tables, we can see that the performance of our proposed system is better than that of the single microphone based methods in Zigel et al. [41] and Popescu and Mahnot [29]. The performance of our proposed system is similar to that of the single microphone method without interference or at low interference level; however, at high interference level, the advantages of our system are evident. Compared with the single microphone based methods, there is an increase of 7.4% TPR for our proposed system (90.63% for our method and 82.81% for the single microphone method) at 75% maximum volume interference level, with 4.6% less FNR being obtained (3.13% for our method and 7.81% for the single microphone method). Besides, it can also be observed that our proposed system is less affected by increasing interference level. For our method, the AUC value drops from 0.9928 with no interference to 0.9738 with 75% maximum volume interference and the geometric mean value drops from 0.9687 to 0.9370. While larger decreases of these two values can be observed from the single microphone based methods in Zigel et al. [41] and Popescu and Mahnot [29], the AUC value drops from 0.9928 with no interference to 0.9293 and the geometric mean value drops from 0.9687 to 0.8737 as the interference level increases to be 75% maximum volume.

5. Discussions

This paper proposes an efficient unsupervised acoustic fall detection system with interference suppression. It inherits the advantages of traditional acoustic sensor based fall detection systems: there is no need for the elderly person to wear special equipment as in the accelerometer based methods; besides, compared with the computer vision based method, there is no problem with sudden illumination change for this proposed acoustic fall detection system and privacy intervening issues are much reduced. Compared with the state-of-the-art acoustic fall detection systems as proposed in Zigel et al. [41], Li et al. [16], Mungamuru and Aarabi [27], Li et al. [15] and Popescu and Mahnot [29], the proposed fall detection system has the following advantages:

1. The proposed fall detection system adopts an unsupervised scheme, which makes use of the features extracted from the normal sound samples constructing an OCSVM model to distinguish falls from non-falls. For the unsupervised method, there is no need for the falling sound samples for model construction compared with the supervised schemes in Zigel et al. [41], Li et al. [16], Mungamuru and Aarabi [27] and Li et al. [15] which need multiple volunteers to simulate extra falling activities to obtain falling sound samples, which is inconvenient; besides, the simulated fall activities from volunteers are different from the real falls from elderly persons.

2. Compared with the unsupervised fall detection system proposed in Popescu and Mahnot [29], this system applies an elegant SS technique for interference suppression by using only two microphones, which makes our proposed fall detection system less sensitive to interferences, as presented in the Experimental section.

Although the source separation method used for interference suppression does reasonably well in the current scenario; as realistic home environments could be more hostile and the level of reverberation very high, the reflections received at the sensors could be very strong and may appear as new sources causing estimation errors. Future work could therefore consider robust techniques for these realistic environments that provide an accurate estimate of the number of sources in the mixture and source separation. Besides, currently the running time on a 2.20 GHz Intel Core i5 processor was approximately 50 s to separate two sound sources from a 3-s i.e. 48,000 sample mixture using a τ grid with 61 values and 16 EM iterations. The computational costs of the source separation method need to be reduced in order to meet the real time demand.

6. Conclusion

In this paper, we proposed a novel unsupervised fall detection system. Acoustic signals (footstep sounds) were collected from an elderly person's normal activities, which were then processed by an efficient dereverberation and source separation technique if the estimated sound number was larger than one; the position information of the sound sources was also exploited in the processing. MFCC features were extracted from the processed acoustic signals and used to construct an OCSVM model to distinguish fall from non-fall sound samples. The performance of this proposed system was evaluated in a simulated environment and better performance was achieved compared with the state-of-the-art single microphone based methods.

Acknowledgments

The authors would like to thank the anonymous reviewers for improving the clarity of our paper. M.S. Khan acknowledges the financial support of UET, Peshawar, the Higher Education Commission (HEC) of Pakistan and Engineering and Physical Sciences Research Council (EPSRC), Grant no. EP/K014307/1.

References

[1] J.B. Allen, D.A. Berkley, Image method for efficiently simulating small-room acoustics, J. Acoust. Soc. Am. 65 (4) (1979) 943–950.

[2] E. Auvinet, F. Multon, A. Saint-Arnaud, J. Rousseau, J. Meunier, Fall detection with multiple cameras: an occlusion-resistant method based on 3-d silhouette vertical distribution, IEEE Trans. Inf. Technol. Biomed. 15 (2) (2011) 290–300.

[3] C. Bishop, Pattern Recognition and Machine Learning, Springer, Berlin, 2006.

[4] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, Cambridge, 2004.

[5] G. Carone, D. Costello, Can Europe afford to grow old? Int. Monet. Fund Finance Dev. Mag. 43 (3) (2006) 28.

[6] M. Estudillo-Valderrama, L. Roa, J. Reina-Tosina, D. Naranjo-Hernandez, Design and implementation of a distributed fall detection system-personal server, IEEE Trans. Inf. Technol. Biomed. 13 (6) (2009) 874–881.

[7] T. Ganchev, N. Fakotakis, G. Kokkinakis, Comparative evaluation of various MFCC implementations on the speaker verification task, in: Tenth International Conference Speech and Computer, Patras, Greece, 2005.

[8] S. Haykin, Adaptive Filter Theory, Prentice-Hall, Upper Saddle River, NJ, 2001.

[9] J. Hsieh, Y. Hsu, H. Liao, C. Chen, Video-based human movement analysis and its application to surveillance systems, IEEE Trans. Multimed. 10 (3) (2008) 372–384.

[10] D. Karantonis, M. Narayanan, M. Mathie, N. Lovell, B. Celler, Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring, IEEE Trans. Inf. Technol. Biomed. 10 (1) (2006) 156–167.

[11] M.S. Khan, S.M. Naqvi, A.u. Rehman, W. Wang, J.A. Chambers, Video-aided model-based source separation in real reverberant rooms, IEEE Trans. Audio Speech Lang. Process. 21 (9) (2013) 1900–1912.

[12] M.S. Khan, S.M. Naqvi, J.A. Chambers, A new cascaded spectral subtraction approach for binaural speech dereverberation and its application in source separation, in: IEEE International Conference on Acoustics Speech and Signal Processing, Vancouver, Canada, 2013.

[13] T. Kim, H.T. Attias, S. Lee, T. Lee, Blind source separation exploiting higher-order frequency dependencies, IEEE Trans. Audio Speech Lang. Process. 15 (1) (2007) 70–79.

[14] K. Lebart, J.M. Boucher, P.N. Denbigh, A new method based on spectral subtraction for speech dereverberation, Acta Acust. United Acust. 87 (3) (2001) 359–366.

[15] Y. Li, K. Ho, M. Popescu, A microphone array system for automatic fall detection, IEEE Trans. Biomed. Eng. 59 (2) (2012) 1291–1301.

[16] Y. Li, Z. Zeng, M. Popescu, K. Ho, Acoustic fall detection using a circular microphone array, in: 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Buenos Aires, Argentina, 2010.

[17] W. Liu, D. Tao, Multiview Hessian regularization for image annotation, IEEE Trans. Image Process. 22 (7) (2013) 2676–2687.

[18] W. Liu, D. Tao, J. Cheng, Y. Tang, Multiview hessian discriminative sparse coding for image annotation, Comput. Vis. Image Underst. 118 (1) (2014) 50–60.

[19] B. Loesch, B. Yang, Source number estimation and clustering for underdetermined blind source separation, in: Proceedings of the IWAENC, Seattle, WA, USA, 2008.

[20] P. Loizou, Speech Enhancement: Theory and Practice, CRC, Boca Raton, FL, 2007.

[21] M. Kangas, A. Konttila, P.L.P.W., T. Jamsa, Comparison of low-complexity fall detection algorithms for body attached accelerometers. Gait Posture 28(2) (2008), 285–291.

[22] H.K. Maganti, D. Gatica-Perez, I. McCowan, Speech enhancement and recognition in meetings with an audio–visual sensor array, IEEE Trans. Audio Speech Lang. Process. 15 (8) (2007) 2257–2269.

[23] M.I. Mandel, R.J. Weiss, D. Ellis, Model-based expectation–maximization source separation and localization, IEEE Trans. Audio Speech Lang. Process. 18 (2) (2010) 382–394.

[24] S. Miaou, F. Shih, C. Huang, A smart vision-based human fall detection system for telehealth applications, in: The Third International Conference on Telehealth (IASTED), Anaheim, CA, USA, 2007.

[25] S. Miaou, P. Sung, C. Huang, A customized human fall detection system using omni-camera images and personal information, in: First Transdisciplinary Conference on Distributed Diagnosis and Home Healthcare, Arlington, VA, USA, 2006.

[26] B. Mirmahboub, S. Samavi, N. Karimi, S. Shirani, Automatic monocular system for human fall detection based on variations in Silhouette area, IEEE Trans. Biomed. Eng. 60 (2) (2013) 427–436.

[27] B. Mungamuru, P. Aarabi, Enhanced sound localization, IEEE Trans. Syst. Man Cybernet.—Part B 34 (3) (2004) 1526–1540.

[28] S.M. Naqvi, W. Wang, M.S. Khan, M. Barnard, J.A. Chambers, Multimodal (audio–visual) source separation exploiting multi-speaker tracking, robust beamforming, and time–frequency masking, IET Signal Process. (Special Issue on Multi-Sensor Signal Processing for Defence: Detection, Localisation and Classification) 6 (5) (2012) 466–477.

[29] M. Popescu, A. Mahnot, Acoustic fall detection using one-class classifiers, in: 31st Annual International Conference of the IEEE EMBS, Minneapolis, MN, USA, 2009.

[30] J. Proakis, D. Manolakis, Digital Signal Processing, 4th Edition, Prentice Hall, New Jersey, 2006.

[31] C. Rougier, J. Meunier, A. St-Arnaud, J. Rousseau, Fall detection from human shape and motion history using video surveillance, in: Twenty-First International Conference on Advanced Information Networking and Applications Workshops (AINAW), Niagara Falls, Ontario, Canada, 2007.

[32] H. Sawada, S. Araki, S. Makino, A two-stage frequency-domain blind source separation method for underdetermined convolutive mixtures, in: Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, 2007.

[33] B. Scholkopf, J. Platt, J. Taylor, A. Smola, R. Williamson, Estimating the support of a high-dimensional distribution, Neural Comput. 13 (7) (2001) 1443–1471.

[34] D. Tax, One-class classification (Ph.D. thesis), TU Delft, NL, 2001.
[35] O. Yilmaz, S. Rickard, Blind separation of speech mixtures via time–frequency masking, IEEE Trans. Signal. Proc. 52 (7) (2004) 1830–1847.
[36] J. Yu, Y. Rui, Y. Tang, D. Tao, High order distance based multiview stochastic learning in image classification, IEEE Trans. Cybern. http://dx.doi.org/10.1109/TCYB.2014.2307862 (in press).
[37] J. Yu, D. Tao, J. Li, J. Chen, Semantic preserving distance metric learning and applications, Inform. Sci. 281 (2014) 674–686, http://dx.doi.org/10.1016/j.ins.2014.01.025.
[38] J. Yu, D. Tao, M. Wang, Adaptive hypergraph learning and its application in image classification, IEEE Trans. Image Process. 21 (7) (2012) 3262–3272.
[39] M. Yu, A. Rhuma, S. Naqvi, J. Chambers, L. Wang, Posture recognition based fall detection system for monitoring an elderly person in a smart home environment, IEEE Trans. Inf. Technol. Biomed. 16 (6) (2012) 1274–1286.
[40] D. Zhou, J. Huang, B. Scholkopf, Learning with hypergraphs: clustering, classification and embedding, in: Proceedings of the Neural Information Processing Systems (NIPS), Vancouver, BC, Canada, 2006.
[41] Y. Zigel, D. Litvak, 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. 56 (12) (2009) 2858–2867.