Efficient Source Separation Algorithms for Acoustic Fall Detection Using a Microsoft Kinect

Yun Li, *Student Member, IEEE*, K. C. Ho*, *Fellow, IEEE*, and Mihail Popescu, *Senior Member, IEEE*

***Abstract*—Falls have become a common health problem among older adults. In previous study, we proposed an acoustic fall detection system (acoustic FADE) that employed a microphone array and beamforming to provide automatic fall detection. However, the previous acoustic FADE had difficulties in detecting the fall signal in environments where interference comes from the fall direction, the number of interferences exceeds FADE's ability to handle or a fall is occluded. To address these issues, in this paper, we propose two blind source separation (BSS) methods for extracting the fall signal out of the interferences to improve the fall classification task. We first propose the single-channel BSS by using nonnegative matrix factorization (NMF) to automatically decompose the mixture into a linear combination of several basis components. Based on the distinct patterns of the bases of falls, we identify them efficiently and then construct the interference free fall signal. Next, we extend the single-channel BSS to the multichannel case through a joint NMF over all channels followed by a delay-and-sum beamformer for additional ambient noise reduction. In our experiments, we used the Microsoft Kinect to collect the acoustic data in real-home environments. The results show that in environments with high interference and background noise levels, the fall detection performance is significantly improved using the proposed BSS approaches.**

***Index Terms*—Blind source separation (BSS), fall detection, microphone array, Microsoft Kinect, nonnegative matrix factorization (NMF).**

I. Introduction

As older live longer and more independent lives, falls have become an important health problem. A recent report [1] shows that one-third of older adults aged above 65 fall each year in the United States. The direct cost for medical care of the fall-related health problems has reached to $23.6 billion in 2005 [1]. Falls have become the leading cause of injury death among older adults [1]. The risk of death among older adults rises significantly when falls fail to be reported to caregivers in time [2]. It has been reported that the longer the lie on the floor, the poorer is the outcome of the medical intervention [3]. To reduce the intervention time, systems that automatically detect and report a fall in a timely manner using different types of sensors have been widely investigated [4]–[8]. Among these sensors, the acoustic sensors are of our great interest due to their low cost and their ability to function in occluded or dark environments.

The use of acoustic sensors for fall detection has been previously reported in the literature [8]–[11]. In [11], we proposed an acoustic fall detection system (acoustic FADE) based on a circular array of microphones. The circular microphone array was able to provide a good estimation of source position under challenging acoustic environment and reduce the number of false alarms significantly using the height estimate of the sound source. In addition, the utilization of a conventional delay-and-sum beamformer (DSB) enhances the source signal so that better fall recognition rate is achieved.

Although the proposed acoustic FADE [11] achieves encouraging fall detection performance, deployment of the device may be difficult and could be invasive in real-home environment due to the large size of the microphone array (see the device pictures in [10], [11]). In addition, the previous acoustic FADE [11] had difficulties in locating the fall signal source using the conventional direction of arrival (DOA) estimator or the sound source localization (SSL) when multiple interferences were present. To address these issues, in [12], we proposed an acoustic FADE based on Microsoft Kinect that consists of a nonuniformly spaced linear array of four microphones. The Kinect system is very compact and has already been deployed in TigerPlace, our living laboratory situated in Columbia, Missouri. Furthermore, the acoustic source position can be deduced accurately from the depth sensing camera [13], [14]. The accurate position aides the use of a more sophisticated adaptive beamformer, such as the minimum variance distortion-less response (MVDR) beamformer, to reduce the interferences coming from multiple directions [15], [16].

Although this Kinect-based approach does better address the multi-interference issues, limitations remain in some particular cases. The first case is when a fall signal comes from the same or closer direction as the interference (e.g., radio, TV). The second case is when a fall occurs outside the field of view (FOV) of Kinect sensors. MVDR in these two cases results in little enhancement and even distorted the source (fall) signal.

To reduce the interferences regardless of the above limitations, blind source separation (BSS) techniques which separate the source of interest out of interferences are considered. Nonnegative matrix factorization (NMF) is a powerful tool [17]

Manuscript received July 12, 2013; revised October 1, 2013; accepted October 24, 2013. Date of publication November 5, 2013; date of current version February 14, 2014. This work was supported in part by the National Science Foundation under Grant CNS-0931607. *Asterisk indicates corresponding author.*

Y. Li is with the Department of Electrical and Computer Engineering, University of Missouri, Columbia, MO 65211 USA (e-mail: yl874@mail.missouri.edu).

*K. C. Ho is with the Department of Electrical and Computer Engineering, University of Missouri, Columbia, MO 65211 USA (e-mail: hod@missouri.edu).

M. Popescu is with the Department of Health Management and Informatics, University of Missouri, Columbia, MO 65211 USA (e-mail: popescum@missouri.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBME.2013.2288783

that is utilized widely in BSS applications. NMF factorizes the mixture signal properly and extracts the bases corresponding to different sources. Traditionally, most of the NMF-based BSS approaches focus on separating the bases of the source of interest using supervised-learning methods such as support vector machine, the Gaussian mixture model, and expectation maximization [18]–[20]. To apply these approaches for fall detection, a training model which consists of prefactorized fall bases needs to be built, making the entire FADE inefficient and impractical. In addition, the previous studies on the multichannel extension of the NMF-based BSS approaches [20], [21] mainly focus on deriving the multiplicative update rules of NMF by maximizing the likelihood of all channels. However, they fail to take the phase information (relative time delays) from multiple channels into account. It is known that the phase information from multiple channels is useful for reducing background noise by utilizing the beamforming techniques [11]. Therefore, to increase the performance and improve the robustness against background noise in the traditional NMF-based BSS approaches [18]–[21], additional noise reduction functionality is developed by exploiting the phase information in multichannel NMF processing.

In this paper, we first propose an efficient single-channel NMF-based BSS approach for separating the fall signal through the analysis of their basis patterns. Then we extend the work for multichannel reception. Finally, we use the TigerPlace [22] dataset collected by Kinect devices to evaluate the fall detection performance using the proposed BSS approaches.

The proposed NMF technique to improve fall detection is generic and is not limited to Kinect platform. It can be applied to other acoustic sensing system when beamforming fails to cancel interferences. They include the interferences arriving at directions close to the fall signal, the interferences are too strong that DOA estimation is not able to locate the fall signal, or the number of interferences exceeds the degrees of freedom of a microphone array.

The structure of the paper is as follow. In Section II, we describe the methodologies of the proposed efficient BSS approaches for separating the fall signal from interferences. Section III provides the description of the experimental data from the Kinect platform and the performance evaluation of fall detection. In Section IV, we present the experimental results and their analysis. Section V gives discussions, conclusions, and future work.

II. Methodologies of the Proposed Efficient BSS Approaches for Separating Fall Signal

A. Single-Channel NMF-Based BSS

NMF [17] is a linear decomposition method that factorizes a nonnegative $\boldsymbol{V}$ matrix (whose elements are not negative) into the product of two nonnegative matrices $\boldsymbol{W}$ and $\boldsymbol{H}$. In mathematical form,

$$\boldsymbol{V} \approx \boldsymbol{W}\boldsymbol{H} \tag{1}$$

where $\boldsymbol{V}$ is M by N, $\boldsymbol{W}$ is M by K, and $\boldsymbol{H}$ is K by N. The factorization expresses each column of $\boldsymbol{V}$ as a linear combination of the columns of $\boldsymbol{W}$, where the coefficients of combination are arranged in a column of $\boldsymbol{H}$. For NMF to be effective, the number of basis vectors K in $\boldsymbol{W}$ should be chosen such that $MK + KN \ll MN$.

NMF is not unique because the factorization cannot be exact unless under very limited special cases. Depending on different measures in the goodness of factorization, different pair of $\boldsymbol{W}$ and $\boldsymbol{H}$ could result. We shall use the Kullback–Leibler (KL) divergence measure in this work since it is found appropriate for music/audio applications [23]. The KL distance measure for NMF is

$$\mathrm{Err}\left(\boldsymbol{V}, \boldsymbol{W}\boldsymbol{H}\right) = \sum_{m=1}^{M}\sum_{n=1}^{N} d_{KL}\left(\boldsymbol{V}_{mn}, [\boldsymbol{W}\boldsymbol{H}]_{mn}\right) \tag{2}$$

$$d_{KL}\left(\boldsymbol{V}_{mn}, [\boldsymbol{W}\boldsymbol{H}]_{mn}\right) = \boldsymbol{V}_{mn}\ln\frac{\boldsymbol{V}_{mn}}{[\boldsymbol{W}\boldsymbol{H}]_{mn}} - \boldsymbol{V}_{mn} + [\boldsymbol{W}\boldsymbol{H}]_{mn}. \tag{3}$$

In (2) and (3), $\boldsymbol{V}_{mn}$ stands for the (m,n)th element of $\boldsymbol{V}$. The definition of $[\boldsymbol{W}\boldsymbol{H}]_{mn}$ is similar and ln is the natural logarithm.

Through taking the derivatives of (3) with respect to matrices $\boldsymbol{V}$ and $\boldsymbol{H}$, the decomposition solution is obtained by iterating the multiplicative updates:

$$\boldsymbol{W}_{mk,l+1} = \frac{\boldsymbol{W}_{mk,l}}{\sum_{n=1}^{N}\boldsymbol{H}_{kn,l}}\sum_{n=1}^{N}\frac{\boldsymbol{V}_{mn}}{[\boldsymbol{W}\boldsymbol{H}]_{mn,l}}\boldsymbol{H}_{kn,l} \tag{4}$$

$$\boldsymbol{H}_{kn,l+1} = \frac{\boldsymbol{H}_{kn,l}}{\sum_{m=1}^{M}\boldsymbol{W}_{mk,l}}\sum_{m=1}^{M}\frac{\boldsymbol{V}_{mn}}{[\boldsymbol{W}\boldsymbol{H}]_{mn,l}}\boldsymbol{W}_{mk,l} \tag{5}$$

in which $k = 1, 2, \ldots, K$ and l represents the iteration index. Iteration stops when the decomposition error is sufficiently small or the maximum number of iterations $T_{\max}$ is reached, where $T_{\max}$ is set to 200 throughout the experiments in this paper.

We now describe how to apply NMF for BSS. The received single-channel mixture signal is of the form

$$s(t) = s_1(t) + s_2(t) + \cdots + s_P(t) \quad t = 0, 1, \ldots, L-1 \tag{6}$$

where the background noise is ignored here for ease of description and it will be taken into account in Section II-C. There are P sources and their components are $s_1(t), s_2(t), \ldots, s_P(t)$. The discrete time index is t and L is the number of data samples. We would like to recover $s_p(t), p = 1, 2, \ldots, P$ from $s(t)$.

The first step to apply NMF is to segment $s(t)$ into a sequence of frames. Let the frame size be M and the amount of overlap be Q. The total number of frames, which is the number of columns of the matrix $\boldsymbol{V}$ to be factorized, is

$$N = [(L-M)/(M-Q)] + 1. \tag{7}$$

In (7), $[(\cdot)]$ represents the largest integer not greater than the real number $(\cdot)$. Each column of $\boldsymbol{V}$ is the fast Fourier transform (FFT) magnitude taken over a frame. Thus, $\boldsymbol{V}$ is simply the magnitude of short-time Fourier transform (STFT) of the mixture signal.

After factorization of such $\boldsymbol{V}$ using KL-NMF, the resultant columns of $\boldsymbol{W}$, $\boldsymbol{w}_k$, are interpreted as the frequency bases of

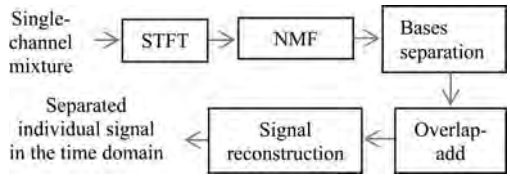

Fig. 1. Processing prototype of single-channel NMF-based BSS approach.

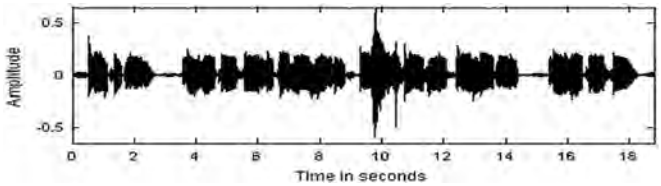

Fig. 2. Segment of mixture signal which contains both a fall and a TV audio signal.

magnitude responses. The kth row of $\boldsymbol{H}$, $\boldsymbol{h}_k^T$, called the temporal basis here, gives the contributions of $\boldsymbol{w}_k$ at different time frames.

In the context of BSS, each source has different frequency bases such that the columns of $\boldsymbol{W}$ can be delineated into P groups, each corresponds to a different source. Let the submatrix $\boldsymbol{W}_p$ be the collection of K_p frequency bases that are classified as from the pth sound source. Thus,

$$\boldsymbol{W}_p = \left[\boldsymbol{w}_{(1),p}, \boldsymbol{w}_{(2),p}, \ldots, \boldsymbol{w}_{(Kp),p}\right] \text{ with } K = \sum_{p=1}^{P} K_p \quad (8)$$

where $\boldsymbol{w}_{(k),p}, k = 1, 2, \ldots, K_p$, is the rearranged kth frequency basis of source p. We collect the rows of $\boldsymbol{H}$ for source p as follows:

$$\boldsymbol{H}_p = \left[\boldsymbol{h}_{(1),p}, \boldsymbol{h}_{(2),p}, \ldots, \boldsymbol{h}_{(Kp),p}\right]^T. \quad (9)$$

The STFT magnitude of source p, $\left|\hat{\boldsymbol{S}}_p\right|$ is reconstructed by

$$\left|\hat{\boldsymbol{S}}_p\right| = \sum_{k=1}^{Kp} \boldsymbol{w}_{(k),p}\boldsymbol{h}_{(k),p}^T = \boldsymbol{W}_p\boldsymbol{H}_p. \quad (10)$$

Since the characteristics of sound source are dominated by its magnitude response, the phase in the STFT of sound source p can be approximated by the phase of the mixture signal $\angle\boldsymbol{S}(m,n)$. Thus, we have

$$\hat{\boldsymbol{S}}_p(m,n) = \left[\boldsymbol{W}_p\boldsymbol{H}_p\right]_{mn} \angle\boldsymbol{S}(m,n). \quad (11)$$

The separated individual source signal in the time domain is constructed by applying the inverse FFT (IFFT) on each frame of $\hat{\boldsymbol{S}}_p$ and joining the time-domain signal in each frame with the overlap-add method [23]. The processing block diagram of the single-channel NMF-based BSS approach is shown in Fig. 1. Recall that FFT magnitude is symmetric, we can reduce the number of rows by nearly a factor of two when doing NMF, hence saving computation time while maintaining the same factorization results.

B. Bases Separation Using the Proposed Unsupervised Learning Method for Single-Channel NMF-Based BSS

The performance of BSS using NMF depends very much on how good we are able to classify each basis vectors in $\boldsymbol{W}$ to one of the P sources. The classification can be made in supervised methods which requires training or in unsupervised manner. We shall propose, in this work, a simple and effective unsupervised technique for the classification of bases vectors.

We use an example to illustrate the rationale and concept of the proposed technique. Let us begin with an 18.8-s segment of acoustic data recorded in the real-world home environment. The segment, denoted as *demoSS*, contains the mixture of one fall signal and one TV audio signal, whose time domain waveform is plotted in Fig. 2.

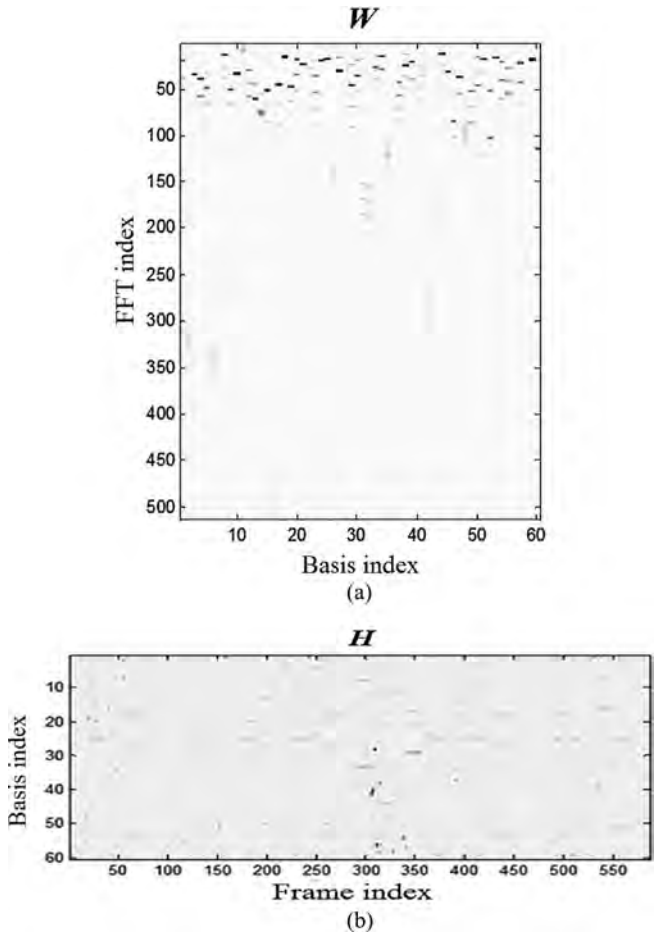

Fig. 3. Plots of KL-NMF decomposed $\boldsymbol{W}$ (a) and $\boldsymbol{H}$ (b) on the data *demoSS*. The lower intensity values (darker) indicate higher magnitude of frequency bases in $\boldsymbol{W}$ and larger temporal factors in $\boldsymbol{H}$.

We choose a frame size of 1024 points and an overlap ratio of 50% between adjacent frames by following [20]. Such settings are commonly used in audio processing. At F_s = 16-kHz sampling frequency, the total number of frames is $N = 586$ as computed by (7). We keep only the first half of the FFT points for decomposition due to symmetry and the size of $\boldsymbol{V}$ is 513 by 586. To achieve sufficient separation accuracy, the number of bases is chosen as $K = 60$ based on some analysis of the experimental data collected at home environments. Fig. 3 gives the NMF decomposition results in gray level images in which $\boldsymbol{W}$ is 513 by 60 and $\boldsymbol{H}$ is 60 by 586.

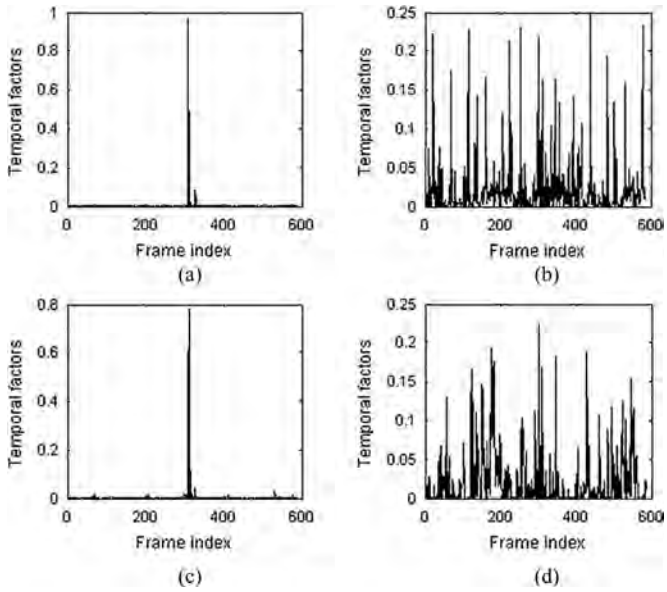

Fig. 4. Plots of selected normalized temporal bases $\tilde{\boldsymbol{h}}_k^T$, (a) $k = 2$, (b) $k = 11$, (c) $k = 31$, and (d) $k = 29$, on data *demoSS*.

We propose to use $\boldsymbol{H}$ instead of $\boldsymbol{W}$ to separate the fall and TV basis components. Let us use $\tilde{\boldsymbol{h}}_k^T$ to denote the kth row of $\boldsymbol{H}$ normalized by its two norm. Fig. 4 shows four of such normalized temporal bases out of 60, where (a) and (c) correspond to fall, (b) and (d) correspond to nonfall signals.

As can be seem from Fig. 4 that (a) and (c) have a significant impulse-like peak in the region where the fall occurs. However, since the TV signal is continuous in time, (b) and (d) show more continuous and random patterns. Based on the observations, we shall obtain features to capture the spiky behaviors in the normalized temporal basis vectors to separate fall and nonfall signals. Thresholding of the features will next be applied to select the fall bases and signal reconstruction will be followed.

1) Energy Ratio Features: The energy ratio is defined as dividing the energy of temporal factors during the occurrence of a fall by the total energy of the entire normalized temporal basis which is known to be one. The energy ratio of temporal basis k can be formulated by

$$E_k = \sum_{n=1}^{\Delta n_L + \Delta n_R} \tilde{h}_k \left(N_0 - \Delta n_L + n\right)^2 \tag{12}$$

where N_0 is the expected frame index when fall occurs, Δn_L and Δn_R are offsets from left and right of N_0 so that a portion for calculating the energy is formed.

In practice, the fall portion in a particular processing window is not known. In sequential processing, BSS is applied to a block of data defined by a sliding window, and a new sequence of data is processed when a sliding window advances to the next. It is reasonable to make the assumption that the fall portion always resides in the very end of a sliding window. To make this happen in our case, the overlap between adjacent processing windows needs to be determined carefully so that the portion covers an entire fall signal. Therefore, we determine the size of the portion as 1 s based on some analysis and the overlapping rate is then determined by $(L - F_s)/L$ in which L refers to the size of a sliding window, which is the same as the number of data samples for BSS as indicated from (6) in Section II-A. Thus, we can rewrite (12) as follows:

$$E_k = \sum_{n=1}^{n_p} \widetilde{h_k} \left(N_0 + n\right)^2 \tag{13}$$

where $n_p = 30$ as computed by (7) and N_0 is the frame when fall occurs.

Depending on the sensing platform, side information may be available to assist in determining the time location where possible fall occurs. For instance, the Kinect platform has depth image sensor that provides very reliable instances when potential fall activities occur.

2) Temporal Bases Separation by Thresholding: To separate the fall temporal bases, the energy ratios of all the temporal bases are passed to a discriminator with a positive threshold Φ. The binary decision rule of the discriminator is expressed by

$$\tilde{\boldsymbol{h}}_k^T \text{ is } \begin{cases} \text{a fall basis} & \text{if } E_k \geq \Phi \\ \text{a non fall basis} & \text{if } E_k < \Phi. \end{cases} \tag{14}$$

Fig. 5 shows the discriminator to separate the fall temporal bases out using thresholding on data *demoSS*.

In Fig. 5, Φ is manually chosen as 0.8. In practice, Φ can be adaptively determined using the noise-plus interference only frames in the data segment before a possible fall signal is detected. As a result, in this example, temporal bases with indices of 2, 3, 8, 13, 15, 16, 31, 33, 35, 55, and 58 are separated out for fall signal reconstruction.

3) Reconstruction of the Separated Source Signal: The selected temporal bases are used for reconstructing the STFT magnitude and the time domain data of the fall signal using the method described in Section II-A. Fig. 6 shows the separated waveform of the fall signal.

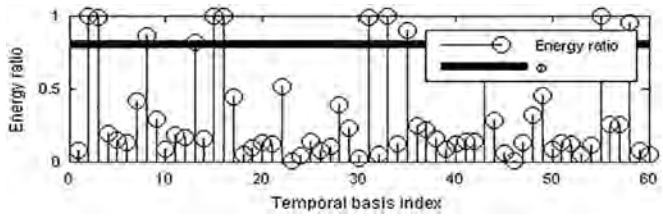

Fig. 5. Separation of fall temporal bases using thresholding on data *demoSS*.

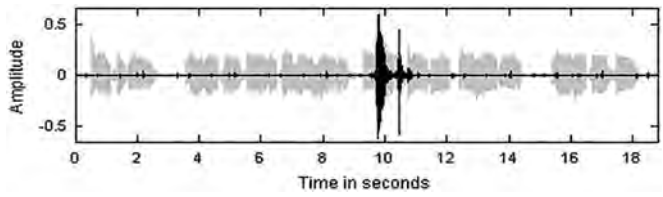

Fig. 6. Separated fall waveform in data *demoSS* using the proposed single-channel NMF-based BSS; original composite signal before source separation (light color), separated fall signal (dark color).

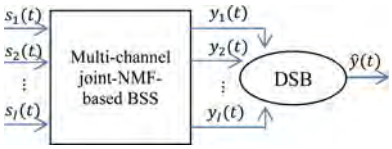

Fig. 7. Processing prototype of multichannel joint-NMF-based BSS approach.

The performance of the proposed fall source separation approach can be evaluated using some standard methods [24]. However, we are more interested in its resulting fall detection performance. The performance evaluation using receiver operating characteristic (ROC) curves [25] will be described in Section III.

C. Proposed Extension to the Multichannel BSS

To apply the proposed single-channel NMF-based BSS to a microphone array, we need to derive its multichannel version. Since NMF does not change the phase of the signals from multiple channels, the conventional DSB can help reduce the background noise of the separated source signal given the source position is well estimated by SSL [26].

Thus, we propose the multichannel joint-NMF-based BSS which is developed by first applying a single NMF to jointly factorize the signals from all channels together. The proposed bases separation approach described in Section II-B is used to separate the individual source signals. DSB is then utilized on the intended signal source to enhance the result. Fig. 7 shows the processing prototype of the proposed multichannel joint-NMF-based BSS approach with J channels.

The motivation for doing NMF jointly over all channels together is to improve factorization accuracy. The signal components in different channels are the same except from different delays and random noises. Hence, in ideal case without noise, the NMF of the signal from a channel is the same as from another channel. This is not the case when noise appears. Joint NMF is able to suppress the noise to achieve better factorization accuracy since the noise components from different channels are independent.

To perform the joint decomposition, we define the matrix $\boldsymbol{V}^{(j)}$ which is formed by the STFT magnitude of the jth channel signal $s_j\,(t)$ by

$$\boldsymbol{V}^{(j)} = \left[\boldsymbol{v}_1^{(j)}, \boldsymbol{v}_2^{(j)}, \ldots, \boldsymbol{v}_N^{(j)}\right] \tag{15}$$

where N is the number of frames as defined in Section II-A. Then we construct the matrix $\boldsymbol{V}$ for factorization as follows:

$$\boldsymbol{V} = \Big[\boldsymbol{v}_1^{(1)}, \boldsymbol{v}_1^{(2)}, \ldots, \boldsymbol{v}_1^{(J)} \vdots \quad \boldsymbol{v}_2^{(1)}, \boldsymbol{v}_2^{(2)}, \ldots, \boldsymbol{v}_2^{(J)} \vdots \quad \cdots\cdots \vdots \quad \boldsymbol{v}_N^{(1)}, \boldsymbol{v}_N^{(2)}, \ldots, \boldsymbol{v}_N^{(J)}\Big]. \tag{16}$$

The number of columns in $\boldsymbol{V}$ is now NJ. After NMF, $\boldsymbol{W}$ remains to have a dimension of $M \times K$ and $\boldsymbol{H}$ has a larger dimension of $K \times NJ$. Then we apply the proposed bases separation in Section II-B to jointly separate the fall bases from $\boldsymbol{H}$. The columns of the matrix $\boldsymbol{H}$ can be partitioned into J submatrices by collecting the columns of $\boldsymbol{H}$ at a regular interval of J. Let $\boldsymbol{h}'_n, n = 1, 2, \ldots, NJ$, be the nth column of $\boldsymbol{H}$. Then

$$\boldsymbol{H}^{(j)} = \left[\boldsymbol{h}'_j, \boldsymbol{h}'_{J+j}, \ldots, \boldsymbol{h}'_{(N-1)J+j}\right], j = 1, 2, \ldots, J \tag{17}$$

is the temporal basis matrix for the data of the jth channel. The number of basis vectors K is set the same as in the single-channel NMF decomposition since each channel encounters the same signal and interferences but shifted by different amounts of relative delays. Therefore, we can reconstruct the separated source signal $y_j\,(t)$ for the jth channel using the selected bases from $\boldsymbol{H}^{(j)}$.

Let $\hat{\boldsymbol{u}}$ be the estimated source position, the enhanced signal $\hat{y}\,(t)$ is obtained by applying DSB to the J separated signals as follows:

$$\hat{y}\,(t) = \frac{1}{J}\sum_{j=0}^{J-1} y_j\,(t + \tau_j\,(\hat{\boldsymbol{u}})) \tag{18}$$

where $\tau_j\,(\hat{\boldsymbol{u}})$ represents the relative correction of delayed samples at the jth channel with respect to $\tau_0\,(\hat{\boldsymbol{u}})$. Note that $\tau_j\,(\hat{\boldsymbol{u}})$ is often not an integer. Efficient implementation of DSB with non-integer time shift is done directly in the frequency domain [11] after BSS before the time-domain conversion.

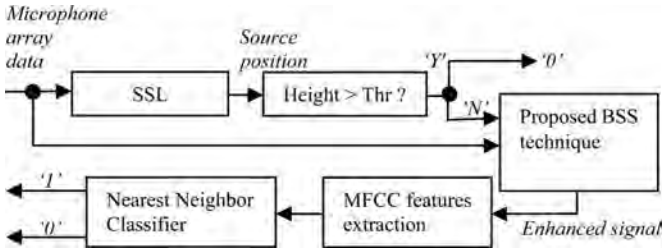

Fig. 8. Processing prototype of the acoustic FADE using the proposed multichannel BSS approach (Thr = positive scalar height threshold, *"Y"*—Yes, *"N"*–No, *"1"*—classified as a fall, and *"0"*—classified as a nonfall).

D. Proposed Acoustic FADE Algorithm

The acoustic FADE algorithm is shown in Fig. 8 which is similar to the previous version in [11] except that the proposed multichannel NMF-based BSS approach is utilized for enhancing the source signal. Interested readers can refer to [11] for the description of SSL, height discriminator, and fall recognition components.

III. Experimental Data and Performance Evaluation

A. Description of Experimental Data From Elderly Homes

The dataset, denoted as D0 used for training in this experiment was collected at TigerPlace in year 2012. The collection of experimental data has been approved by the Institutional Review Board at the University of Missouri. The data collection was performed in ten apartments. In each apartment, a Kinect device is installed above the apartment door viewing the entire living room. Each month, a stunt actor (see [11] for the three stunt actors' information) comes into each room and performs falls and other activities under the instructions of

TABLE I
DESCRIPTION OF FALL TYPES PERFORMED BY STUNT ACTORS

Initial status	Motivations	Directions
Standing	Loose balance or momentary loss of consciousness	Forward, backward, left, right or straight down.
Sitting	From the chair	Forward, left, right, sliding forward out of the chair as the chair slides back or sliding backward out of the chair as it slides front.
Lying	From the bed or couch	Fall to side, upper body falls first or hips and shoulders fall first.
Walking	Tripping and slipping	Trip forward or sideways, slip forward, sideways or backward.

the nursing staff [27]. D0 consists of 120 segments of clean (very low noise) fall signals and 120 segments of clear non-fall signals and they are extracted from the Kinect recordings. Each segment has duration of 1 s. The types of falls are categorized based on the faller's initial status: standing, sitting, lying, tripping, and slipping. The details of the fall types are described in Table I.

The types of nonfalls are partially the same as the one used in [11] including phone ringing, talking, typing, and some challenging ones such as dropping books, dropping boxes, and rolling bottles. Segments of typical TV audio and background noises are also included.

Another dataset, named D1 is a three-day long sequence which was collected recently in TigerPlace using the living room Kinect. D1 has a total of 50 falls performed by the stunt actors and also contains a variety of typical daily activities from the resident (the resident is living alone). The stunt actors A and B are different from the ones for creating the dataset D0. The detailed data and stunt actors' information are described in Tables II and III, respectively.

Typically, the source positions or DOAs are estimated using SSL algorithms such as steered response power [26] and multiple signal classification [28]. One advantage of using the Kinect platform is that the DOA of the signal of interest can be obtained with reasonable accuracy from the depth images, thereby reducing the complexity in performing DOA estimation.

B. Simulation of the Mixture Signals at Kinect Microphones

A linear microphone array, such as the one in a Kinect device, can utilize the azimuth DOA only. Let θ represents the DOA of the faller and ϕ_{p_I} represents the DOA of the p_I^{th} interference out of P_I. Under the assumption of far-field acoustic model, the source configuration for a Kinect microphone array is drawn in Fig. 9.

In this experiment, we use one 10-s segment of prerecorded TV audio signals to simulate the interferences. To mix the fall signals with the interferences, the duration of each fall signal in D0 is extended to 10 s by padding zeros in front of the fall signal. Thus, the fall signal resides in the tail portion as assumed in Section II-B.

TABLE II
DESCRIPTION OF THE THREE-DAY LONG DATASET (KINECT CAPTURES FROM 7 AM TO 11 PM EACH DAY AND TOTAL 48 HOURS)

		Description of the general activites
Day1 (stunt actor A, room1)	7 AM —	Cooking*, watching TV*, door opening and closeing*, phone ringing, talking*, etc.
	10:17 AM	Stunt actors and nursing staffs come, talking, 4 falls out of Kinect FOV (2 occluded by a coach and 2 in the kitchen) with TV on, 8 falls very close to a TV with TV on followed by 4 falls with TV off.
	10:40 AM	Stunt actors and nursing staffs left.
	— 11 PM	Cooking*, watching TV*, bathing*, etc.
Day2 (stunt actor B, room2)	7 AM —	Cooking*, resident left*, etc.
	10:53 AM	Stunt actors and nursing staffs come, talking, 4 falls very close to a TV with TV on, 8 falls with TV on, 8 falls out of Kinect FOV (4 in the kitchen and 4 occluded by furnitures) with TV on followed by 4 falls with TV off.
	11:28 AM	Stunt actors and nursing staffs left.
	— 11 PM	Resident comes back*, watching TV*, bathing*, etc.
Day3 (stunt actor B, room3)	7 AM —	Cooking*, watching TV*, phone ringing, talking*, reading*, etc.
	11:06 AM	Stunt actors and nursing staffs come, talking, 10 falls out of Kinect FOV (5 in the kitchen and 5 in the bedroom) with TV on.
	11:40 AM	Stunt actors and nursing staffs left.
	— 11 PM	Cooking*, watching TV*, bathing*, etc.
Comments		All stunt actor falls performed in this dataset consists of all types as described in Table I. The TV volumn changes at each fall (-5dB<SIR<5dB). SNR is below 5dB during the data collection. Multiple-interference cases such as multiple people are talkings while a fall occurs are included. No resident falls are reported during the absence of stunt actors and nursing staffs. '*' represents the activity types identified by listening only.

TABLE III
STUNT ACTORS' INFORMATION

	Gender	Age	Height	Weight
Stunt actor A	Male	31	5'5"	225 lbs
Stunt actor B	Female	39	5'3"	130 lbs

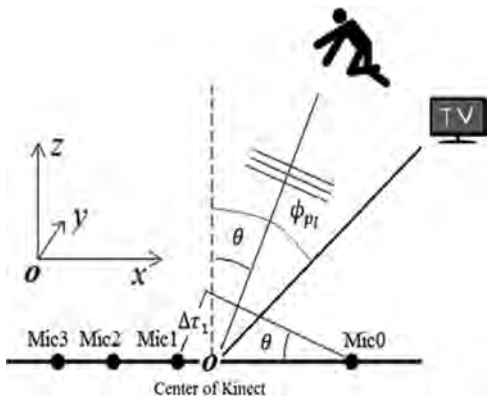

Fig. 9. Source configuration of a Kinect microphone array. "Mic" denotes the Kinect microphone.

Without loss of generality, we design the experiments particularly for the case which is illustrated in Section I as the motivation of using BSS approaches. The case describes a single interference whose DOA is the same as that of a fall signal source, i.e., with the specific settings: $\phi_1 = \phi_{\mathrm{TV}} = \theta, P_I = 1$.

Thus, the procedures of generating the mixture signal at each Kinect microphone are described as follows:

1) Make Mic0 marked in Fig. 9 as reference channel at which the mixture signal is generated by simply adding the zero-padded fall signal to the interference signal. Specify the DOA of ith fall signal, $\theta^{(i)}, i = 1, 2 \ldots, 120$, randomly (following uniform distribution) bounded by $\left|\theta^{(i)}\right| \leq 57^{\circ}$ (horizontal angular FOV) [29]. Use the segment of TV audio as the interference source and scale its amplitude based on the specified signal-to-interference ratio (SIR) and signal-to-noise ratio (SNR) levels. Then specify its DOA corresponding to the ith fall signal by $\phi_{\mathrm{TV}}^{(i)} = \theta^{(i)}$.
2) Correct the phase of the i^{th} fall signal at Mic1, Mic2, Mic3, and Mic4 according to $\theta^{(i)}$. Similarly, adjust the phase of the corresponding interference signal corresponding $\phi_{\mathrm{TV}}^{(i)}$.
3) Add the adjusted ith fall signal to the corresponding interferences at the four mics to generate the mixture signals. Store the ith 4-channel mixture signal in a new dataset, named as D2.

To compute the STFT magnitude of the 10-s mixture signal, we choose the frame size as 1024 points with 512 overlapped points. The number of frames is 311 as computed by (7). We set the number of bases to $K = 50$. Thus, the factorized matrix $\boldsymbol{W}$ has a dimension of 513 by 50 and matrix $\boldsymbol{H}$ has a dimension of 50 by 311.

C. Performance Evaluations Using Both Simulated and Real-World Data

To validate the improvement of the proposed BSS approaches, we use D0 to create 10 classifiers by decomposing D0 into ten folds in a cross-validation manner. For each classifier, the testing fold is replaced by the corresponding data from D2 in which interference and noise are added. The procedure of generating a cross-validation ROC curve, determination of ROC operating threshold, and the definitions of performance metric indices such as sensitivity, specificity, accuracy, and area under the ROC curve (AUC) are all explicitly described in [11].

To evaluate the performance of the proposed BSS approaches on real-world dataset D1 (see Section III-A), we use D0 as training dataset and detect falls in D1 using the proposed sequential processing as described in Section II-B) at a particular classifier threshold. Ground truths of fall time locations are obtained based on the Kinect RGB video [29]. Thus, the number of detected falls and the number of false alarms per hour are both obtained. By evaluating D1 at different classifier thresholds, an ROC curve is then generated.

IV. Experimental Results and Analysis

The purpose of the experiments is to validate the proposed source separation approaches for enhancing the fall signal by reducing both the interferences and the background noise. The fall detection performances using both approaches—single-channel NMF-based BSS denoted by SglSS and multichannel joint-NMF-based BSS denoted by MulSS are evaluated when SIR and SNR are at different levels. We also evaluate the performance using the mixture signal from a single channel denoted by Mix as the benchmark.

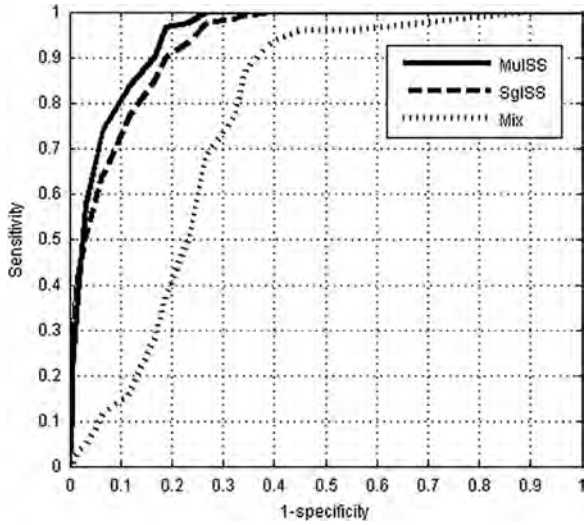

Fig. 10. Tenfold cross-validation ROC curves of fall detection using MulSS, SglSS, and Mix when SNR = "clean" and SIR = −5 dB.

TABLE IV
TENFOLD CROSS-VALIDATION RESULTS OF MULSS, SGLSS, AND MIX AT DIFFERENT SNR LEVELS AND SIR = −5 DB (SENSITIVITY, SPECIFICITY, AND ACCURACY VALUES ARE IN % AT THE OPERATING THRESHOLDS)

		'clean'	5dB	0dB	-5dB	-10dB
MulSS	AUC	0.946	0.941	0.933	0.901	0.844
	Sensitivity	97	96	93	87	81
	Specificity	81	79	80	78	78
	Accuracy	89	88	87	83	80
SglSS	AUC	0.931	0.899	0.853	0.713	0.482
	Sensitivity	90	93	83	68	53
	Specificity	81	70	76	82	61
	Accuracy	86	82	80	75	57
Mix	AUC	0.763	0.726	0.669	0.452	0.246
	Sensitivity	93	87	73	61	23
	Specificity	61	61	60	48	47
	Accuracy	77	74	67	55	35

A. Performance Evaluation Using Simulated Dataset

D2 is used to evaluate the performance. The SNR levels are specified as: "clean," 5, 0, −5, and −10 dB. Channel independent white Gaussian noise is used to simulate the background noise and its amplitude is scaled based on the specified SNR level for each channel. The background noise is added to the data in D2 to obtain the new mixture signal. The cross-validation ROC curves when SNR = "clean" and SIR = −5 dB are shown in Fig. 10. The evaluation metric indices at different SNR levels are tabulated in Table IV.

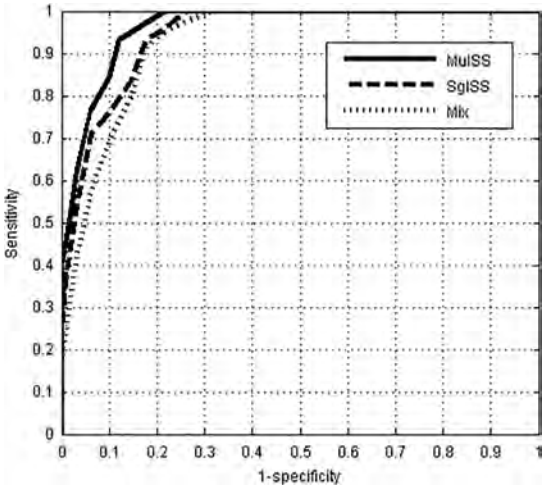

Fig. 11. Tenfold cross-validation ROC curves of fall detection using MulSS, SglSS, and Mix when SNR = "clean" and SIR = 0 dB.

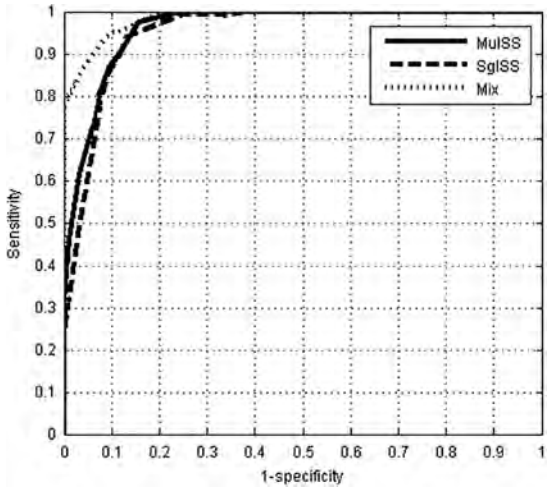

Fig. 12. Tenfold cross-validation ROC curves of fall detection using MulSS, SglSS, and Mix when SNR = "clean" and SIR = 5 dB.

TABLE V
TENFOLD CROSS-VALIDATION RESULTS OF MULSS, SGLSS, AND MIX AT DIFFERENT SNR LEVELS AND SIR = 0 DB (SENSITIVITY, SPECIFICITY, AND ACCURACY VALUES ARE IN % AT THE OPERATING THRESHOLDS)

		'clean'	5dB	0dB	-5dB	-10dB
MulSS	AUC	0.962	0.937	0.921	0.905	0.846
	Sensitivity	93	97	93	83	86
	Specificity	88	76	78	83	73
	Accuracy	91	87	86	83	80
SglSS	AUC	0.944	0.922	0.873	0.856	0.578
	Sensitivity	93	89	86	84	57
	Specificity	82	78	75	73	65
	Accuracy	88	84	81	79	61
Mix	AUC	0.927	0.919	0.833	0.688	0.498
	Sensitivity	92	87	84	67	61
	Specificity	82	85	71	65	54
	Accuracy	87	86	78	66	58

TABLE VI
TENFOLD CROSS-VALIDATION RESULTS OF MULSS, SGLSS, AND MIX AT DIFFERENT SNR LEVELS AND SIR = 5 DB (SENSITIVITY, SPECIFICITY, AND ACCURACY VALUES ARE IN % AT THE OPERATING THRESHOLDS)

		'clean'	5dB	0dB	-5dB	-10dB
MulSS	AUC	0.961	0.953	0.946	0.918	0.864
	Sensitivity	98	93	90	91	88
	Specificity	84	83	86	77	75
	Accuracy	91	88	88	84	82
SglSS	AUC	0.951	0.946	0.927	0.847	0.664
	Sensitivity	94	92	89	75	61
	Specificity	87	83	83	85	77
	Accuracy	91	88	86	80	69
Mix	AUC	0.984	0.943	0.859	0.776	0.589
	Sensitivity	95	84	91	86	58
	Specificity	90	94	69	66	66
	Accuracy	93	89	80	76	62

Fig. 10 and Table IV show that the proposed multichannel BSS approach achieves the best among the three at each SNR level, which indicates that the beamforming has positive effect even the fall and interference signals come from the same direction. The proposed multichannel BSS approach is still able to achieve up to 80% accuracy at the most challenging acoustics (SIR = −5 dB and SNR = −10 dB).

Fig. 11 shows the ROC curves when SNR = "clean" and SIR = 0 dB. The evaluation metric indices at different SNR levels are tabulated in Table V.

Fig. 11 and Table V show similar conclusion to the previous one when SIR = 0 dB that the proposed multichannel BSS approach achieves the best results at each SNR. Since the interference level is smaller, the overall performance over the three approaches is slightly better than the one when SIR = −5 dB.

Fig. 12 shows the ROC curves when SNR = "clean" and SIR = 5 dB. The evaluation metric indices at different SNR levels are tabulated in Table VI.

Fig. 12 and Table VI show that when both the interference and background noise are sufficiently small, the performance of a single-channel mixture improves much more than the other two proposed approaches. This indicates that in such high SNR and SIR acoustic environment, a mixture signal should be used for fall detection instead of using the interference-removed signals from the proposed BSS approaches. This is because the negative effect of BSS in removing some small fall components out-weighs the gain in eliminating the interference, where the classifier is able to handle the interference in the case when it is not too strong.

B. Real-World Dataset Processing Strategy

In order to optimize the overall fall detection performance on a sequence of real-world dataset, the following three issues need to be addressed.

1) Occlusion of a Fall: When a fall is occluded and out of sensor FOV, it is impossible for the fall signal to propagate through direct paths to the acoustic sensors, which results in the failure of the proposed multichannel BSS approach. The proposed single-channel BSS approach, however, is able to address the occlusion issue. Since a Kinect is used, the simultaneous

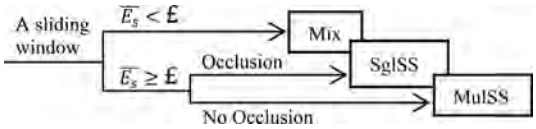

Fig. 13. Block diagram of the processing strategy combining all three evaluation approaches.

occurrence of the absence of the resident (out of living room) and the occlusion of a potential fall (occluded by objects) can be detected easily through depth sensing (no depth data is generated when occlusion occurs).

2) Number of Interferences Exceeding the Degrees of Freedom of a Microphone Array: The proposed BSS approaches separate the sources based on the temporal and spectral information of the mixture composite. Therefore, they are not affected by the number of interferences and their DOAs. As a result, multiple interference $(P_I \geq 2)$ situations can be treated by BSS as a single-interference case $(P_I = 1)$ through lumping all interferences together. The real-world experiment studies include multiple interference cases, as described in the comments of Table II, to validate the performance of the proposed BSS approaches.

3) Performance Decline in High SNR and SIR Levels: As indicated in Fig. 12, the performance of the proposed BSS approaches in such a situation are not as good as the one using a single-channel mixture. To detect such acoustic environment with extremely high SNR and SIR levels, the overall level of background noise and interference is measured by calculating the average energy in the portion of the data which may contain only the noise and interference signal. In a particular sliding window as described in Section II-B1, we are able to calculate the average energy $\overline{E_s}$ of the portion of length $L-F_s$ over J channels by

$$\overline{E_s} = \frac{1}{J\left(L - Fs\right)} \sum_{j=1}^{J} \sum_{t=0}^{L-Fs-1} s_j\left(t\right)^2 . \tag{19}$$

The decision of utilizing a single-channel mixture for processing is made by comparing $\overline{E_s}$ with a threshold £ which is premeasured in the home environment during the absence of the resident or the midnight when the resident is asleep.

Fig. 13 shows the block diagram of the processing strategy which combines all three detectors (MulSS, SglSS, and Mix) as described in Section IV-A in order to optimize the overall fall detection performance. In the proposed acoustic FADE shown in Fig. 8, the "proposed BSS technique" block is replaced by the following processing strategy.

C. Performance Evaluation Using Real-World Dataset

The performance using the real-world dataset D1 is evaluated using three detectors—proposed acoustic FADE using the processing strategy described in Section IV-B, the acoustic FADE (replacing the block "proposed BSS technique" in Fig. 8 by a conventional DSB) proposed in [11] and a single-channel mixture (Mix). The corresponding ROC curves are plotted in Fig. 14.

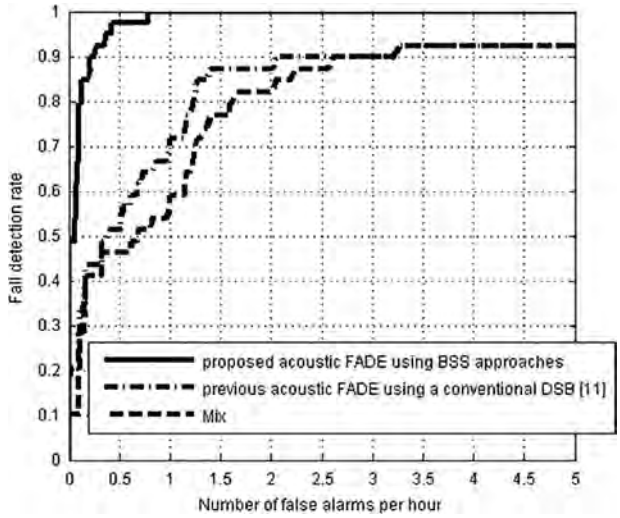

Fig. 14. ROC curves of fall detection using the strategy combining all three evaluation approaches and a single-channel mixture on the real-world dataset.

The energy threshold £ described in Section IV-B is chosen as 0.003 in this particular apartment.

Fig. 14 shows significant improvement using the proposed strategy combining all three fall detection approaches compared with the one using only a single-channel mixture. The proposed system is able to achieve about 0.4 false alarms per hour when the fall detection rate reaches up to 98% or only one fall is missing. The performance using single-channel mixture however, never achieves 100% detection rate until its false alarm rate reaches to about 200/hour (not shown in Fig. 14 for better comparison). The performance using the previous acoustic FADE [11] is slightly better than that of Mix but is still uncompetitive with the proposed one. The significant improvement of the proposed acoustic FADE indicates that the proposed BSS approaches are able to remove the interference at challenging cases such as fall signals come from closer to the interference or being occluded. The other two approaches fail to handle these cases.

V. Discussions and Conclusion

This paper proposes both single-channel and multichannel joint-NMF-based BSS techniques for incorporation into FADE to address the following three issues: 1) when a fall comes from the same or a close direction of interference; 2) the number of interferences exceeds the degrees of freedom of a microphone array; or 3) the occlusion of a fall. These issues result in complete failure of the adaptive beamforming used in the previous work [11], [12].

The proposed bases separation method in the single-channel NMF-based BSS approach is quite effective. It is able to achieve 80% accuracy (see Table IV) when interference is very strong (SIR $= -5$ dB) and background noise is moderate (SNR $=$ 0 dB).

The proposed multichannel NMF-based BSS approach has been illustrated to be more resilient to background noise as compared to the single-channel one. However, the fall position

must be known before using DSB. The fall position can be obtained reliably and efficiently in the Kinect platform. It is able to achieve 80% accuracy when both interference and noise reach to the highest levels in the experiments.

It is also worth to mention that the proposed single-channel and multichannel BSS approaches perform worse than the single-channel mixture when the interference and background noise levels are both sufficiently small (e.g., SNR $>$ 30 dB and SIR $>$ 5 dB) as shown in Fig. 12. As described in Section II-B, the potential fall bases in matrix $\boldsymbol{H}$ are separated based on the energy ratio, which is independent of the changing of the overall basis energy. On the other hand, it has been found that some of the factorized bases are actually shared by both fall signal and noise-plus-interference components, which causes the loss of some spectral information for reconstructing the fall signal. Therefore, even extremely small level of interference and background noise in the mixture will distort the reconstructed fall signal. Without any distortion of the fall signal and with the domination of the fall energy, a single-channel mixture is expected to have better fall detection performance, due to the fact that the classifier is able to handle small amount of interference and noise.

To optimize the overall performance on real dataset, a processing strategy is proposed by incorporating all the approaches. A three-day dataset collected in elder home is used for validating the processing strategy. It has been shown that only one fall is missing at a cost of 0.4 false alarms per hour by utilizing the processing strategy. In conclusion, the proposed source separation approaches appear to have robust behavior to the changing acoustic properties in real-world environments.

Future work will focus on improving the bases separation methods by reducing the potential loss of spectral information for fall signal reconstruction due to the sharing of the bases. In addition, more real-world dataset will be collected and analyzed to further evaluate the fall detection performance of our FADE based on the proposed BSS approaches. Techniques to further reduce the amount false alarms will also be explored.

ACKNOWLEDGMENT

The authors would like to thank the staff in the Center for Eldercare and Rehabilitation Technology, University of Missouri for their efforts on the assistance in data collection.

REFERENCES

[1] Center for Disease Control (CDC), (2013). [Online]. Available: http://www.cdc.gov/HomeandRecreationalSafety/Falls/index.html

[2] R. J. Gurley, N. Lum, M. Sande, B. Lo, and M. H. Katz, "Persons found in their homes helpless or dead," *N. Engl. J. Med.*, vol. 334, no. 26, pp. 1710–1716, 1996.

[3] C. G. Moran, R. T. Wenn, M. Sikand, and A. M. Taylor, "Early mortality after hip fracture: Is delay before surgery important," *J. Bone Joint Surg.*, vol. 87, pp. 483–489, 2005.

[4] 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, Dec. 2009.

[5] D. Anderson, R. H. Luke, J. Keller, M. Skubic, M. Rantz, and M. Aud, "Linguistic summarization of activities from video for fall detection using voxel person and fuzzy logic," *Comput. Vision Image Understanding*, vol. 113, no. 1, pp. 80–89, Jan. 2009.

[6] C. Rougier, J. Meunier, A. St-Arnaud, and J. Russeau, "Fall detection from human shape and motion history using video surveillance," in *Proc. 21st Int. Conf. Adv. Inform. Netw. Appl. Workshops*, 2007, pp. 875–880.

[7] A. Sixsmith, N. Johnson, and R. Whatmore, "Pyrolitic IR sensor arrays for fall detection in the older population," *J. Phys. IV France*, vol. 128, pp. 153–160, 2005.

[8] 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 *Proc. IEEE 30th Annu. Int. Eng. Med. Biol. Soc. Conf.*, Aug. 2008, pp. 4628–4631.

[9] Y. Li, Z. L. Zeng, M. Popescu, and K. C. Ho, "Acoustic fall detection using a circular microphone array," in *Proc. IEEE 32th Annu. Int. Eng. Med. Biol. Soc. Conf.*, Buenos Aires, Argentina, Sep. 2010, pp. 2242–2245.

[10] Y. Li, M. Popescu, and K. C. Ho, "Improving automatic sound-based fall detection using iVAT clustering and GA-based feature selection," in *Proc. IEEE 34th Annu. Int. Eng. Med. Biol. Soc. Conf.*, Aug. 2012, pp. 5867–5870.

[11] Y. Li, K. C. Ho, and M. Popescu, "A microphone array system for automatic fall detection," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 2, pp. 1291–1301, May 2012.

[12] Y. Li, T. Banerjee, M. Popescu, and M. Skubic, "Improvement of acoustic fall detection using Kinect depth sensing," in *Proc. IEEE 35th Annu. Int. Eng. Med. Biol. Soc. Conf.*, Jul. 2012, pp. 6736–6739.

[13] E. E. Stone and M. Skubic, "Capturing habitual, in-home gait parameter trends using an inexpensive depth camera," in *Proc. IEEE 34th Annu. Int. Eng. Med. Biol. Soc. Conf.*, Aug. 2012, pp. 5106–5109.

[14] T. Banerjee, J. M. Keller, and M. Skubic, "Resident identification using Kinect depth image data and Fuzzy clustering techniques," in *Proc. IEEE 34th Annu. Int. Eng. Med. Biol. Soc. Conf.*, Aug. 2012, pp. 5102–5105.

[15] H. Van Trees, *Optimum Array Processing*. New York, USA: Wiley, 2002.

[16] P. Sinha, A. D. George, and K. Kim, "Parallel algorithm for robust broadband MVDR beamforming," *J. Comput. Acoust.*, vol. 10, no. 1, pp. 69–96, 2002.

[17] D. D. Lee and H. S. Seung, "Learning the parts of objects with nonnegative matrix factorization," *Nature*, vol. 401, pp. 788–791, 1999.

[18] M. Helén and T. Virtanen, "Separation of drums from polyphonic music using nonnegtive matrix factorization and support vector machine," in *Proc. Eur. Signal Process. Conf.*, Istanbul, Turkey, 2005.

[19] E. M. Grais and H. Erdogan, "Regularized nonnegative matrix factorization using Gaussian mixture priors for supervised single channel source separation," *Comput. Speech Lang.*, vol. 27, no. 3, pp. 746–762, May 2013.

[20] A. Ozerov and C. Févotte, "Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 3, pp. 550–563, Mar. 2010.

[21] H. Sawada, H. Kameoka, S. Araki, and N. Ueda, "Efficient algorithms for multichannel extensions of Itakura–Saito nonnegative matrix factorization," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, 2012, pp. 261–264.

[22] TigerPlace, an assisted living facility in Columbia, MO, USA. (2007). [Online]. Available: http://eldertech.missouri.edu/overview.htm

[23] C. Févotte, N. Bertin, and J. Durrieu, "Nonnegative matrix factorization with the Itakura–Saito divergence: With application to music analysis," *Neural Comput.*, vol. 21, pp. 793–830, 2009.

[24] E. Vincent, R. Gribonval, and C. Fevotte, "Performance measurement in blind audio source separation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 4, pp. 1462–1469, Jul. 2006.

[25] M. H. Zweig and G. Campbell, "Receiver-operating characterisitc (ROC) plots: A fundamental evaluation tool in clinical medicine," *Clin. Chem.*, vol. 39, no. 8, pp. 561–577, 1993.

[26] J. P. Dmochowski, J. Benesty, and S. Affes, "A generalized steered response power method for computationally viable source localization," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 8, pp. 2510–2526, Nov. 2007.

[27] M. Rantz, M. Aud, G. Alexander, B. Wakefield, M. Skubic, R. H. Luke, D. Anderson, and J. Keller, "Falls, technology, and stunt actors: New approaches to fall detection and fall risk assessment," *J. Nursing Care Quality*, vol. 23, no. 3, pp. 195–201, 2008.

[28] C. Kwan, K. C. Ho, G. Mei, Y. Li, Z. Ren, R. Xu, Y. Zhang, D. Lao, M. Stevenson, V. Stanford, and C. Rochet, "An automated acoustic system to monitor and classify birds," *EURASIP J. Applied Signal Process.*, vol. 2006, pp. 1–19, 2006.

[29] (2013). [Online]. Available: http://msdn.microsoft.com/en-us/library/jj131033.aspx

Yun Li (S'08) was born in China. He received the B.S. degree in electrical engineering from the Harbin Engineering University, Harbin, China, in 2007, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Missouri, Columbia, MO, USA, in 2009 and 2013, respectively.

He is currently a Postdoctoral Fellow in the Center for Engineering and Health, Northwestern University, Evanston, IL, USA. His research interests include array signal processing, source localization, detection and estimation, blind source separation, machine learning and developing applications for eldercare.

K. C. Ho (M'91–SM'00–F'09) was born in Hong Kong. He received the B.Sc. degree (First Class Hons.) in electronics and the Ph.D. degree in electronic engineering, both from the Chinese University of Hong Kong, Hong Kong, in 1988 and 1991, respectively.

He was a Research Associate in the Royal Military College of Canada from 1991 to 1994. He joined the Bell-Northern Research, Montreal, Canada, in 1995, as a Member of Scientific Staff. He was a Faculty Member in the Department of Electrical Engineering, University of Saskatchewan, Saskatoon, Canada, from September 1996 to August 1997. Since September 1997, he has been with the University of Missouri and is currently a Professor in the Electrical and Computer Engineering Department. He is the Associate Rapporteur of ITU-T Q16/SG16: Speech Enhancement Functions in Signal Processing Network Equipment, and was the Rapporteur of ITU-T Q15/SG16: Voice Gateway Signal Processing Functions and Circuit Multiplication Equipment/Systems from 2009 to 2012. He is the Editor of the ITU-T Standard Recommendations G.160: Voice Enhancement Devices and G.168: Digital Network Echo Cancellers. His research interests include sensor array processing, source localization, detection and estimation, wireless communications, and the development of efficient signal processing algorithms for various applications.

Dr. Ho is the Chair of the Sensor Array and Multichannel Technical Committee in the IEEE Signal Processing Society. He was an Associate Editor of the IEEE Transactions on Signal Processing from 2003 to 2006 and from 2009 to 2013, and the IEEE Signal Processing Letters from 2004 to 2008. He received the Junior Faculty Research Award in 2003 and the Senior Faculty Research Award in 2009 from the College of Engineering, University of Missouri, Columbia, MO, USA. He is an inventor of 20 patents in the United States, Europe, Asia, and Canada on geolocation and signal processing for mobile communications.

Mihail Popescu (SM'08) received the M.S. degree in medical physics, the M.S. degree in electrical engineering, and the Ph.D. degree in computer science all from the University of Missouri, Columbia, MO, USA, in 1995, 1997, and 2003, respectively.

He is currently an Associate Professor with the Department of Health Management and Informatics, University of Missouri. His research interests include eldercare technologies, fuzzy logic, and ontological pattern recognition.