34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Improving automatic sound-based fall detection using iVAT clustering and GA-based feature selection

Yun Li, Mihail Popescu, *senior IEEE member,* K. C. Ho, *IEEE fellow*

Abstract — Falls represent an important health problem for older adults. This issue continues to generate interest in the research and development of fall detection systems. In previous work we proposed an acoustic fall detection system (acoustic-FADE) that employs an 8-microphone circular array to automatically detect falls. Acoustic-FADE has achieved encouraging results: 100% detection at 3% false alarm rate in laboratory tests. In this paper, we use a dataset from previous work to investigate how to further improve AFADE performance. To analyze the relationship between fall and non-fall signatures we used the improved visual assessment of tendency (iVAT) clustering algorithm in conjunction with a nearest neighbor based distance to find the most challenging false alarms. Then, we employed a genetic algorithm (GA) framework to perform feature selection and find the mel-frequency cepstral coefficients (MFCC) that improve the classification performance. We found that using only three MFCC coefficients (1, 28, 29) instead of our previous choice (1,2,3,4,5,6) improves the classification performance.

I. Introduction

Falls represent an important health problem for older adults. One in every three adults over 65 fall each year and acquire moderate to severe injuries, such as head traumas and hip fractures, that can increase the risk of early death [1-2]. In the mean time, less than half of the older adults who fall report the issues to the health caregivers [1]. If an older adult who lives alone falls onto the floor and is not able to ask for assistance in a short period of time, he or she is more likely to suffer from hypothermia, dehydration, pressure sores or rhabdomyolosis [3]. The large number of unreported falls can greatly raise the chances of causing more serious health problems due to the delay of the medical intervention. Therefore, it is imperative to develop a system that can effectively detect a fall as soon as it occurs so that an immediate assistance can be provided.

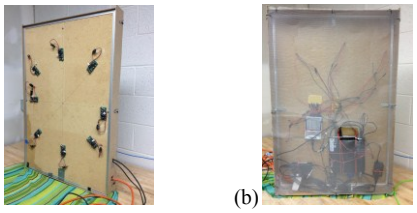
To address the intervention delay problem, multiple fall detection methods have been investigated in the past several years. The fall detection systems previously reported consist of two types: wearable and non-wearable devices. Most wearable devices use accelerometers to detect a fall by measuring the applied acceleration along the vertical axis [4]. The main advantages of the wearable devices are that they are inexpensive and they can be deployed both indoor and outdoor. Their main disadvantages are that they can't be worn during night and they might be rejected by older adults. Many non-wearable devices, such as floor vibration sensors [5], video cameras [6], infrared cameras [7], bed sensors [8], radar sensors [9] and acoustic sensors [10-13] have been investigated recently. Of all the non-wearable devices, the acoustic sensors have the following advantages: low cost, wide "field of view" and night time effectiveness.

In our latest work [13] we proposed an acoustic fall detection system, acoustic-FADE, which is more robust to background noises and reverberation effects compared to its previous versions [10-12]. In some conditions with low background noise such as night time, acoustic-FADE achieves 100% detection at 3% false alarm rate. Although acoustic-FADE performance is encouraging, we are trying to reduce the false alarms further while detecting all the falls.

In this paper, we try to achieve this goal by analyzing the false alarms that "are close" to falls and find ways to remove them. We perform closeness analysis by clustering the fall, non-fall signature dissimilarity matrix using iVAT. Then, we performing feature selection we intend to make these false-alarms disappear, i.e. to be classified as non-falls. For instance, a 'backward fall' and 'dropping of a book' may be difficult to distinguish if the wrong features are used, although they can be easily differentiated by the human ear. To reduce the false alarms, we developed a GA-based feature selection method to make acoustic-FADE more robust to the false alarms and easier to detect falls.

II. Description of Acoustic-FADE

The proposed acoustic-FADE consists of two components – the acoustic sensor hardware and the data processing software. The acoustic sensor hardware consists of 8 microphones uniformly located along a circular with 25cm radius. The picture of the modified version of acoustic-FADE hardware is shown in Fig. 1.



(a) (b)

Fig. 1. (a) Front view of acoustic-FADE hardware. (b) Back view of acoustic-FADE hardware.

Fig. 1 (a) shows 8 microphone mounted on a wood board

Y.L. is with the ECE Dept., University of Missouri (corresponding author, e-mail: yl874@mail.mizzou.edu).

M. P. is with the Health Mngmt. & Inform. Department, University of Missouri, Columbia, MO 65211, USA (email: popescum@missouri.edu)

D.H. is with Elect. and Comp. Eng. Dept., University of Missouri, Columbia, MO 65211, USA (email: hod@missouri.edu).

installed inside a thin box. The processing hardware including a DAQ (Data Acquisition) device and an EeeBox PC is installed inside the box, on the back of the microphone board, as shown in Fig. 1 (b). The sampling frequency of the DAQ is set to 20 KHz and each data sample is quantized to 12 bits.

Details of the data processing algorithm used in acoustic-FADE are not relevant for this paper but the interested reader is referred to [13].

III. EXPERIMENTAL METHODOLOGY

A. Dataset description

We obtained the approval for our fall detection research project from the Institutional Review Board (IRB) of the University of Missouri, Columbia (please refer to http://eldertech.missouri.edu/ for more details). In this paper we use a dataset (see Table I) that was collected in a home-like laboratory room [13].

TABLE I. DESCRIPTION OF THE DATASET

ID	Fall types (name format: 'type' -'trend')	ID	Non-fall types
1	Balance-Forwards	21	Closing window
2	Balance-Backwards	22	Typing keyboard
3	Balance-Left	23	Key shaking
4	Balance-Right	24	Machine noise
5	Lose consciousness-Forwards	25	Phone ringing
6	Lose consciousness-Backwards	26	Knocking door
7	Lose consciousness-Left	27	Talking
8	Lose consciousness-Right	28	Sitting on a bed
9	Trip and fall-Forwards	29	Sitting on a couch
10	Trip and fall-Sideways	30	Sitting on a chair
11	Slip and fall-Forwards	31	Normal walking
12	Slip and fall-Sideways	32	Slow walking
13	Slip and fall-Backwards	33	Shoes shuffling
14	Reach fall (chair)-Forwards	34	Dropping book
15	Reach fall (chair)-Left	35	Dropping tennis ball
16	Reach fall (chair)-Right	36	Dropping metal can
17	Slide fall-Forwards	37	Dropping wood box
18	Slide fall-Backwards	38	Dropping plastic bottle
19	Couch fall-Upper body first	39	Rolling a can
20	Couch fall-Hips first	40	Rolling a plastic bottle

The dataset consists of 120 fall files (20 types, 6 files per type) and 120 non-fall files (20 types, 6 files per type). The falls were performed by 3 well-trained stunt actors instructed by our nursing staff. Many non-falls in the dataset shown in Table I consist in sounds "similar" to a fall (such as dropping an object or sitting hard on a piece of furniture) and were intentionally introduced to challenge the classifier. Note that the non-fall types with ID 21-30 are above-ground sound sources while the ones with ID 31-40 are on-the-ground sound sources. Details about the data collection methodology (how to train stunt actors to fall like elderly, information about the stunt actors, etc) can be found in [13].

B. Dissimilarity matrix calculation

1) MFCC features

We calculate the MFCC features of the enhanced signal (the enhanced signal is described in [13]) of each file in the dataset. The MFCC matrix of the enhanced signal in the p^{th} file has the following form [13]:

$$\boldsymbol{C}_p = \begin{bmatrix} C_{1,1} & \cdots & C_{1,121} \\ \vdots & \ddots & \vdots \\ C_{N,1} & \cdots & C_{N,121} \end{bmatrix} \quad p = 1,2,\dots,240 \tag{1}$$

where the rows represent the MFCC coefficient index and the columns is the sub-frame index. We have previously [13] used N=6 MFCC coefficients. In this study, we will determine N using a GA framework.

2) Dissimilarity matrix

The dissimilarity matrix $\boldsymbol{D}$ calculates the pair-wise distance among all the sounds (falls and non-fall) signatures. The ordering of the sounds in the matrix is the same as in Table I (i.e first 120 sounds represent falls and the last 120 represent non-falls). The dissimilarity matrix is computed as:

$$\boldsymbol{D}_{ij} = \left\| \boldsymbol{C}_i - \boldsymbol{C}_j \right\| \quad , \quad i,j = 1,2,\dots,240 \tag{2}$$

where $\|\cdot\|$ is the Frobenius norm. $\boldsymbol{D}$ is then normalized as $\boldsymbol{D_n}=\boldsymbol{D}/\max(\boldsymbol{D})$ and displayed as an intensity image in which each pixel represents a dissimilarity value.

C. Evaluation of detection effectiveness using iVAT

The visual assessment of cluster tendency (VAT) algorithm [14] is used for determining the cluster tendency or the possible number of clusters in a set of objects, based on visual assessment. First, the objects are suitably reordered. Second the dissimilarity matrix is regenerated based on the new order of the objects. Finally, the reordered dissimilarity matrix $\widetilde{\boldsymbol{D}}$ is normalized and displayed as an intensity image in which the dark blocks of pixels along the diagonal indicate the cluster tendency. The improved VAT (iVAT) algorithm proposed by [15] is used for (harder) cases where VAT fails to indicate any cluster tendency.

In this study, we apply iVAT to the dissimilarity matrix $\boldsymbol{D_n}$ to generate the intensity image so that we can better understand the patterns of the dataset. In addition, the iVAT image can identify which types of falls or non-falls might be difficult to classify due to the closeness in the feature space.

D. GA-based feature selection

It is well known that the performance of the classifier and the computational cost can be greatly improved by feature selection. The GA framework is one of the possible solutions to the feature selection problem. The main idea of a GA framework is to try various feature combinations and choose the one the maximized some objective function (called "fitness function"). Particular solutions to the problem are called individuals or chromosomes. In our case the fitness function attempts to minimize the intra-cluster dissimilarity and maximize the inter-cluster dissimilarity. Suppose we take M MFCCs from N, then the number of possible combinations for the M MFCCs is C_N^M. Then the fitness function of the k^{th}, $k = 1,\dots,C_N^M$, combination of M MFCCs $\boldsymbol{\epsilon}_{k,M} = [\epsilon_{k,M}(1), \epsilon_{k,M}(2), \dots, \epsilon_{k,M}(M)]$ ($\epsilon_{k,M}(m)$ is the index of m^{th} MFCC in the given combination and $\epsilon_{k,M}(m) \in [1, N]$) can be written as

$$Fitness(\boldsymbol{\epsilon}_{k,M}) = \frac{1}{\omega_1\left([\boldsymbol{D}_{f,f}(\boldsymbol{\epsilon}_{k,M})]/[\boldsymbol{D}_{f,nf}(\boldsymbol{\epsilon}_{k,M})]\right)^2 + \omega_2\left([\boldsymbol{D}_{nf,nf}(\boldsymbol{\epsilon}_{k,M})]/[\boldsymbol{D}_{f,nf}(\boldsymbol{\epsilon}_{k,M})]\right)^2} \tag{3}$$

in which $\boldsymbol{D}$ is the dissimilarity matrix defined in III.*B* and $[\boldsymbol{D}] = \frac{1}{120^2}\sum_{i=1}^{120}\sum_{j=1}^{120} D_{ij}$ is the average dissimilarity. $\omega_n, n = 1,2$ is a weighting factor and satisfies $\sum_{n=1}^{2}\omega_n = 1$. The subscript '*f*' and '*nf*' denote a fall and a non-fall, respectively. For instance, $\boldsymbol{D}_{f,f}$ is the fall-fall sub-matrix of $\boldsymbol{D}$ (see Fig. 2).

By encoding the coefficient indices $\boldsymbol{\epsilon}_{k,M}(m)$, *m*=1, 2,…, *M*, in binary format, an individual has the following form:

$$\overbrace{0\,1\,\ldots\,1}^{\epsilon_{k,M}(1)}\ \overbrace{1\,1\,\ldots\,0}^{\epsilon_{k,M}(2)}\cdots\ \overbrace{0\,1\,\ldots\,0}^{\epsilon_{k,M}(M)}$$

The GA runs by reproducing the 'fittest' individuals from the previous generation and terminates when the fitness doesn't change significantly. For each *M,* we obtain the best fitness of all generations and the fittest overall individual $\boldsymbol{\epsilon}_{\hat{k},M}$ (the solution).

The most interesting part about this feature selection procedure is that it is independent of the algorithm employed for classification (SVM, neural networks, Bayes, etc). This procedure strictly reflects the quality of the features. We believe that decoupling the two problems (feature and classifier selection) is a better way to analyze a classification problem.

IV. Results and Discussions

A. (Fall, non-fall) dissimilarity matrix

The normalized gray intensity image of the dissimilarity matrix described in III.*B* is shown in Fig. 2.

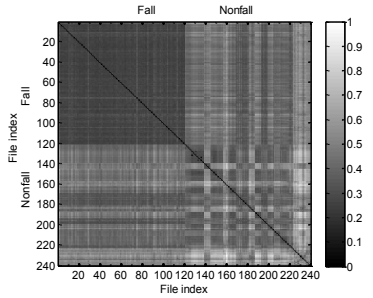


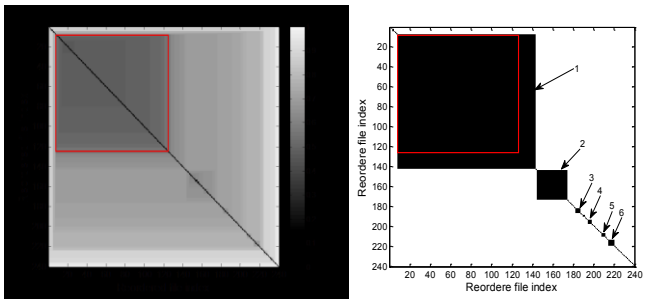
Fig. 2. Normalized gray intensity image of the dissimilarity matrix.

As we see from Fig. 2, the darker block at the upper left indicates that falls have very low dissimilarities with each other. On the other hand, non-falls (bottom right block) have partially higher dissimilarities among themselves due to the variety of activities involved in false alarm production. The other two blocks (nonfall-fall and fall-nonfall) indicate higher dissimilarities between falls and non-falls although a few exception may be noted.

B. iVAT clustering

The image of $\widetilde{\boldsymbol{D}}$, the iVAT-ed $\boldsymbol{D_n}$ matrix, is shown in Fig. 3 (a). In Fig. 3 (a), the large darker block included in the red square box represents the cluster which includes all falls, which means falls are more similar to themselves than to non-falls. There are several small darker blocks along the diagonal which may indicate several non-fall clusters. To have a better understanding of these small clusters, we need to further process the iVAT image.

This could be done by converting the iVAT gray image into a binary image based on a properly selected threshold. We determine the threshold based on Otsu's method which tries to separate two classes of intensity values by minimizing their intra-class variances [16]. The resulting Otsu threshold was 0.67. The binary iVAT image is shown in Fig. 3 (b).



(a) (b)

Fig. 3 (a). Normalized gray intensity iVAT image (The darker block inside the red square box represents the cluster including all falls).
(b). Binary iVAT image based on the selected threshold (The possible clusters are marked by the numbers. The region inside the red square box represents all falls in cluster 1 and the outer region of the red square box represents nonfalls in cluster 1).

In Fig. 3 (b), the 6 dark blocks along the diagonal clearly indicate 6 possible clusters in the dataset. Clusters 2~6 consist of non-falls and cluster 1 consists of both falls and non-falls. The non-falls present in cluster 1 are the challenging one since they are more similar to falls than to non-falls for the given threshold. The cluster number and the included non-fall types (type ID, see Table I) are tabulated in Table II.

TABLE II. CLUSTERS WITH RESPECT TO THEIR INCLUDED NON-FALL TYPES (THE MOST CHALLENGING NON-FALL TYPE IDS ARE HIGHLIGHTED).

Cluster#	1	2	3	4	5	6
Type IDs	**28,29,30,33**	22,23,25, 27,32	31	33	35	36, 37

In Table II we found that the four most challenging non-fall types in the given dataset are: "sitting on a bed", "sitting on a couch", "sitting on a chair" and "shoe shuffling" (accentuated walking). The non-fall types found in the other 5 clusters are very dissimilar to falls. The non-fall types not found in any of the clusters have more diversity in their feature patterns; however, they are dissimilar to falls since they have high dissimilarities to falls in the non-block regions in the iVAT image (see Fig. 3 (a)).

C. GA-based feature selection

In the GA framework we set the the initial population size to 500, the crossover probability to 0.7, the mutation probability to 0.03 and the weighting factors $\omega_1 = 0.9, \omega_2 = 0.1$. We chose *M*=1,2,…,6 coefficients out of 30 coefficients and run GA for each *M*. The best fitness, the indices of the corresponding MFCCs and the challenging non-fall types found by their iVATs for each *M* are tabulated in Table III.

As we see in Table III, the best fitness is obtained for M=3. Consequently, the best choice of coefficients should be $\boldsymbol{\epsilon}_{\hat{k},3}$ (1th, 28th and 29th MFCC), which improves the fitness by more than 50% from 1.84 obtained using the original model (1st to 6th MFCCs). It is worth noting that one of the challenging

non-fall types, "sitting on a chair", is missing in the new model.

TABLE III. THE GA-BASED FEATURE SELECTION RESULTS (THE BEST CHOICE IS HIGHLIGHTED).

M	1	2	**3**	4	5	6
$\epsilon_{\hat{k},M}$	1	1,27	**1,28,29**	1,26, 28,29	1,17, 23.27.29	1,19, 26,27, 29,30
Best fitness	2.71	2.74	**2.78**	2.72	2.67	2.62
Challenging type IDs	28,29, 30,33	28,29, 33	**28,29, 33**	28,29, 30,33	28,29, 30,33	28,29, 30,33

In addition, we found that the first and the last several MFCCs (27~30th) are significant in the selected features. These observations are in agreement to what we have found in previous work [13] that lower MFCC are important in fall discrimination (low frequency, 10-200 Hz) while higher MFCC are important in non-falls (high frequency) discrimination. What we did not know in [13], and we found here, was the particular MFCC coefficients needed to achieve best discrimination.

D. Fall detection performance evaluation

To evaluate the improvement of fall detection using the MFCC features selected in IV.*C*, we generate the 10-fold cross-validation ROC curves (height discriminator is not included in the recognition) [13] for the given dataset in both cases: old and new features (see Fig. 4).

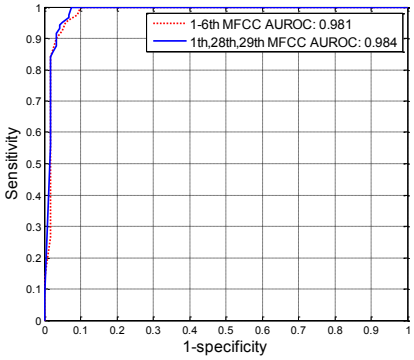


Fig. 4. Comparison of 10-fold cross-validation ROC curves in both cases of using the unselected MFCCs (1~6th) and selected MFCCs (1th, 28th and 29th).

As we see from Fig. 4, fall detection performance improves slightly (0.3%) when the new features are used. In addition, when both cases reach 100% sensitivity, the 1-specificity or false alarm rate of using new features is reduced by about 3%.

V. CONCLUSIONS

This paper presents a study for investigating and improving the effectiveness of the proposed acoustic-FADE. By interpreting the information in the iVAT image, we found that the ability to detect falls is independent of their types since falls are very similar to themselves. We found four types of non-falls that are difficult to distinguish from falls.

We introduced a feature selection method based on a GA framework that is independent of the type of classifier. The GA-based feature selection helps the classifier become more robust to the challenging non-fall types. Under the same experimental conditions, we improved the performance evaluation by 3% using the new features compared with the ones used in the state-of-the-art work in [13].

Although the improved Acoustic-FADE achieved higher performance, further improvement needs to be made to deal with the diversity and uncertainty of the real-world data. We are currently developing a comprehensive fall detection system that uses sensor and classifier fusion to address the challenging types of false alarms that acoustic-FADE alone can't eliminate.

ACKNOWLEDGEMENTS

This work has been supported in part by the NSF grant CNS-0931607.

REFERENCES

[1] Center for Disease Control (2011), http://www.cdc.gov/HomeandRecreationalSafety/Falls/adultfalls.html.

[2] B.H. Alexander, F.P. Rivara, and M.E. Wolf, "The cost and frequency of hospitalization for fall–related injuries in older adults," *American Journal of Public Health*, 82(7), pp. 1020–3, 1992.

[3] P.J. Ratclife, J.G. Ledingham, P. Berman, G.K. Wilcock, and J. Keenean, "Rhabdomyolysis in elderly people after collapse," *British Med. J.*, vol. 288, pp. 1877-8, 1984.

[4] N. Noury, A. Fleury, P. Rumeau, *et al*., "Fall detection-principles and methods," in *Proc. 29th Annu. Int. IEEE EMBS Conf.*, Lyon, France, Aug. 2007, pp. 1663-1666.

[5] M. Alwan, P.J. Rajendran, S. Kell, D. Mack, S. Dalal, M. Wolfe, and R. Felder, "A smart and passive floor-vibration based fall detector for elderly," in *Proc. 2nd IEEE Int. Conf. Inf. & Comm. Tech.*, Damascus, Syria, Apr. 2006, pp. 1003-1007.

[6] 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," *Computer Vision and Image Understanding*, vol. 113 (1), pp. 80-89, Jan. 2009.

[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] J. Hilbe, E. Schulc, B. Linder, and C. Them, "Development and alarm threshold evaluation of a side rail integrated sensor technology for the prevention of falls," *Int. J. Med. Inform.*, vol. 79(3), pp. 173-80, 2010.

[9] L. Liu, M. Popescu, M. Skubic, M. Rantz, T. Yardibi, and P. Cuddihy, "Automatic fall detection based on Doppler radar motion," in *Proc. 5th Int. Conf., Pervasive Computing Technologies for Healthcare*, Dublin, Ireland, May, 2011, pp. 222-225.

[10] M. Popescu, Y. Li, M. Skubic, M. Rantz, "An acoustic fall detector system that uses sound height information to reduce the false alarm rate," in *Proc. 30th Int. IEEE EMBS Conf.*, Vancouver, BC, pp. 4628-4631, Aug. 20-24, 2008.

[11] Y. Li, Z.L. Zeng, M. Popescu, K.C. Ho, "Acoustic fall detection using a circular microphone array," in *Proc. 32th Int. IEEE EMBS Conf.*, Buenos Aires, Argentina, Sept. 1-4, 2010.

[12] Y. Li, M. Popescu, K.C. Ho, and D.P. Nabelek, "Improving acoustic fall recognition by adaptive signal windowing," in *Proc. 33th Int. IEEE EMBS Conf.*, Boston, USA, Aug. 30-Sept.3, 2011, pp. 7589-7592.

[13] Y. Li, K.C. Ho, M. Popescu, "A microphone array system for automatic fall detection," *IEEE Trans. Biomedical Engineering*, vol. 59, pp. 1291-1301, 2012.

[14] J.C. Bezdek and R.J. Hathaway, "VAT: A tool for visual assessment of (cluster) tendency," in *Proc. Int. Joint Conf. Neural Networks*, Honolulu, HI, 2002, pp. 2225–2230.

[15] T.C. Havens and J.C. Bezdek, "An efficient formulation of the improved visual assessment of tendency (iVAT) algorithm," *IEEE Trans. Knowledge and Data Engineering*, vol. 24, pp. 813-822, 2011.

[16] N. Otsu, "A threshold selection method from gray-level histograms," 715 *IEEE Trans. Syst., Man., Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979.