Applied Acoustics 140 (2018) 296–300

Contents lists available at ScienceDirect

Applied Acoustics

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

Technical note

Fall detection through acoustic Local Ternary Patterns

Syed M. Adnan[a], Aun Irtaza[a], Sumair Aziz[b], M. Obaid Ullah[b], Ali Javed[c], Muhammad Tariq Mahmood[d,*]

[a] Department of Computer Science, University of Engineering & Technology (UET), Taxila, Pakistan
[b] Department of Electronics Engineering, University of Engineering & Technology (UET), Taxila, Pakistan
[c] Department of Software Engineering, University of Engineering & Technology (UET), Taxila, Pakistan
[d] School of Computer Science & Engineering, Korea University of Technology and Education, Cheonan, Republic of Korea

ARTICLE INFO

Keywords:
Acoustic-LTP
SVM
Fall detection
Classification

ABSTRACT

In this paper, we propose a framework that detects falls by using acoustic Local Ternary Patterns (acoustic-LTPs) by analyzing environmental sounds. The proposed method suppresses silence zones in sound signals and distinguishes overlapping sounds. Acoustic features are extracted from the Separated source components by using the proposed acoustic-LTPs. Subsequently, fall events are detected through a support vector machine (SVM) based classifier. The performance of the proposed descriptor is evaluated against state-of-the-art methods that are applied on well-known sound databases. A comparative analysis demonstrates that the proposed descriptor is more powerful and reliable in terms of fall detection than other methods, and it also performs well in a multi-class environment. Moreover, the proposed descriptor possesses a rotation invariant property, and therefore, it demonstrates significant resistance against the rotated sound signals.

1. Introduction

Elderly people living alone face distress when they fall and are unable to call for help. In the case of elderly people, a fall may result in life changing injury, severely affecting the quality of life. Moreover, a protracted delay in first aid after a fall further increases the risk of mortality [1,2]. Therefore, early fall detection is crucial to provide timely necessary help, avoid complications, and reduce hospitalization costs.

In the literature, fall detection for elderly people has been proposed using either wearable devices with sensing technologies based on accelerometers or through environmental sensors, i.e., pressure sensors, microphones, video cameras, and floor vibration sensors installed at various locations throughout a building [3–6]. Wearable devices used for fall detection are inconvenient and obtrusive for patients. In [7], a Doppler radar-based fall detection method was proposed to recognize human activity. In [8], fall detection was performed using Radar's effective non-intrusive sensing modality by detecting human motion. In [6], a wavelet transform based method was used to detect human falls using a ceiling mounted Doppler range control radar. The major drawback of using a radar-based Doppler system is their limited applicability. On the other hand, the privacy issues are convoluted in video based methods.

Of the various environmental sensor-based approaches, an acoustic analysis of environmental sounds provides an effective alternative to overcome the drawbacks of both wearable and non-wearable solutions [9,10]. Li, Ho et al. proposed an acoustic analysis for fall detection using the Mel-frequency Cepstral Coefficients (MFCC) features and nearest neighbor (NN) classifier [11]. Shaukat, Ahsan et al. performed daily sound recognition for elderly people using the MFCC, Linear Predictive Coding (LPCs) and non-spectral features [12]. The main drawback of these methods is the selection of many irrelevant features that negatively affect the results of the classification [13]. Another drawback is the inherent complexity that makes the combination less suitable to implement with real time systems. Zigel, Litvak et al. analyzed floor vibration waves and fall sounds in combination for fall detection [13]. Khan, Yu et al. presented a fall detection system using acoustic signals collected from sounds of footsteps [14]. Popescu and Mahnot classified MFCC features through a nearest neighbor (NN), support vector machines (SVM), and Gaussian mixture classifiers for fall detection [15]. The common reason to use MFCCs for fall detection are the lower dimensionality of features [16]. However, during the audio signal acquisition, several environmental factors affect this process and induce noise in the collected sound data. Also, various operating conditions also influence the extracted MFCC features and deteriorate their quality, and these limitations can result in a mismatch when MFCCs are

* Corresponding author.
E-mail addresses: syed.adnan@uettaxila.edu.pk (S.M. Adnan), aun.irtaza@uettaxila.edu.pk (A. Irtaza), sumair.aziz@uettaxila.edu.pk (S. Aziz), Obaid.ullah@uettaxila.edu.pk (M.O. Ullah), ali.javed@uettaxila.edu.pk (A. Javed), tariq@koreatech.ac (M.T. Mahmood).

https://doi.org/10.1016/j.apacoust.2018.06.013
Received 25 March 2018; Received in revised form 11 June 2018; Accepted 14 June 2018

used for classifier training and recognition of a fall event [14]. In addition, MFCC feature extraction is also a computationally complex process and consequently becomes difficult to implement using hardware devices. Different feature extraction techniques can be combined with MFCC to improve the performance by reducing the effects of noise, further increasing the hardware implementation costs. For such reasons, a more effective feature extraction technique needs to be carried out to ensure a better classification performance in fall detection.

In this letter, we propose a novel feature extraction scheme for acoustic signals through acoustic Local Ternary Patterns (acoustic-LTP). The LTP feature descriptors were initially proposed for face recognition [17]. However, such features have never been reported to represent audio signals, which are predominantly 1-D in nature. In addition, the concept of uniform and rotation invariance for audio signals has also been introduced. We emphasize that the rotation invariance is also a fundamental requirement for audio descriptors.

2. Proposed fall detection framework

2.1. Silent zone suppression

A general architecture of a fall detection framework is shown in Fig. 1. In the first step, an input audio signal is processed to suppress the silence zones. When an analog audio signal $y(t)$ is captured from the environment for small intervals of time, it is sampled to obtain a discrete-time signal $y'[n]$ consisting of N' samples. The discrete input signal is divided into F' non-overlapping frames/windows with a fixed length l. Let $q_i \in \Omega_p$, $i = 0, \cdots, 7$, and q_i is the i^{th} neighboring sample in the neighborhood Ω_p centered at p. The discrete audio signal $y'[n]$ has an amalgamation of various audio streams comprising a living environment, including the sound of a fall. The audio stream also contains silence zones which need to be suppressed. By using the HMM model [18] and the FAST-ICA [19], low and high frequency signals are discriminated. The posterior probability for the acoustic events is larger than the posterior probability of the silence period. The frames belonging to the acoustic events, having higher posterior probabilities, are segmented from the sources through the FAST-ICA algorithm. Thus, a source signal $y[n]$ with N samples and F frames is available for further processing.

2.2. Acoustic Local Ternary Patterns

In the second step, the acoustic features of the $y[n]$ signal are extracted through the proposed acoustic Local Ternary Patterns (acoustic-LTP). Acoustic-LTP are locally computed by encoding each frame Ω_p of the audio signal $y[n]$. To compute the ternary pattern, we compute the magnitude difference between the central p and the surrounding samples q_i. Using a threshold t (t = 0.00008) signal values in the range of width $\pm t$ around the central sample p are quantized to zero. Values above $p + t$ are quantized to 1 and below $p - t$ are quantized to -1. Hence, a three-valued function s is given by:

$$s(q_i, p, t) = \begin{cases} +1, & q_i - (p + t) \geqslant 0 \\ 0, & (p + t) < q_i < (p - t) \\ -1, & q_i - (p - t) \leqslant 0 \end{cases} \tag{1}$$

where $s(q_i, p, t)$ represents the acoustic signal using a three-valued ternary pattern. To reduce the number of patterns, they are further split into upper $s_u(.)$ and lower $s_l(.)$ patterns. In the $s_u(.)$ pattern, only $+1$ values are retained while all other values are replaced with zeros.

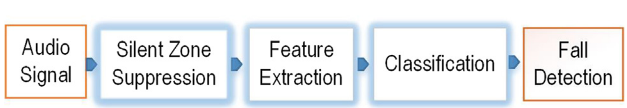

Fig. 1. Architecture of the proposed fall detection framework.

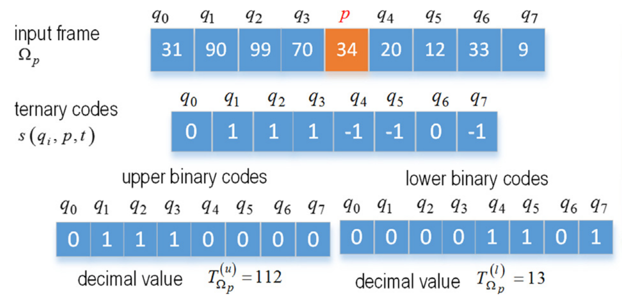

Fig. 2. Acoustic Local Ternary Pattern (acoustic-LTP) Computation.

$$s_u(q_i, p, t) = \begin{cases} 1, & s(q_i, p, t) = +1 \\ 0, & otherwise \end{cases} \tag{2}$$

Similarly, in $s_l(.)$, -1 values are retained as 1 while all other values are replaced with zeros.

$$s_l(q_i, p, t) = \begin{cases} 1, & s(q_i, p, t) = -1 \\ 0, & otherwise \end{cases} \tag{3}$$

The procedure of computing acoustic-TLP is shown in Fig. 2. Uniform patterns are well-known for computer vision applications [17]. They capture most of the attributes of a signal. The ratio uniform patterns is very high as compared to non-uniform patterns. Among the patterns in $s_u(.)$ and $s_l(.)$, the upper uniform patterns $s_u^{uni}(.)$ and lower uniform patterns $s_l^{uni}(.)$ are computed and encoded through their decimal values.

$$T^{(u)}_{\Omega_p} = \sum_{i=0}^{7} s_u^{uni}(q_i, p, t).\ 2^i \tag{4}$$

$$T^{(l)}_{\Omega_p} = \sum_{i=0}^{7} s_l^{uni}(q_i, p, t).\ 2^i \tag{5}$$

For the feature descriptor, two histograms from the upper and lower codes are computed. For each uniform pattern, one bin is assigned and all non-uniform patterns are grouped into a single bin.

$$h_u(k) = \sum_{f=1}^{F} \delta(T_f^{(u)}, k) \tag{6}$$

$$h_l(k) = \sum_{f=1}^{F} \delta(T_f^{(l)}, k) \tag{7}$$

where k denotes the histogram bins corresponding to the uniform acoustic-LTP codes and $\delta(.)$ is the Kronecker delta function.

We observed that the first twenty uniform patterns from the upper and lower patterns are sufficient to capture all variations in the data. Thus, the dimension of the feature vector is two times as long as the dimension of each histogram. The 40-dimentional feature vector $\mathbf{x}$ is formed by concatenating two histograms.

$$\mathbf{x} = [\mathbf{h}_u\ \mathbf{h}_l] \tag{8}$$

2.3. Classification

Finally, fall and non-fall events are classified through a classifier trained using support vector machines (SVM) [20]. For the learning classifier, training data with fall and non-fall audio features with known targets, consisting of M pairs $(\mathbf{x}^{(i)}, y^{(i)})$, $i = 1, \cdots, M$, are prepared where $y^{(i)} \in \{1, -1\}$ specifies the fall and non-fall classes. Hyperplanes linearly separating the two classes are given as,

$$\begin{cases} \mathbf{w}^T\mathbf{x}^{(i)} + b \geqslant 1, & if\ y^{(i)} = 1 \\ \mathbf{w}^T\mathbf{x}^{(i)} + b < 1, & if\ y^{(i)} = -1 \end{cases} \tag{9}$$

where $\mathbf{w}$ is the weighting vector and b is the bias. The objective is to maximize the separation between two planes by minimizing the norm $\|\mathbf{w}\|$ which can be formulated as a quadratic optimization problem.

$$\min_{\mathbf{w}} \|\mathbf{w}\| \text{ s. t. } \quad y(\mathbf{w}^T\mathbf{x}^{(i)} + b) \geqslant 1 \tag{10}$$

The two events can be classified by using the discriminating function $f(\mathbf{x}^{(i)}) = \text{sign}(\mathbf{w}^T\mathbf{x}^{(i)} + b)$ such that if $f(\mathbf{x}^{(i)}) = +1$ then results in fall event otherwise non-fall event.

3. Experiments and results

3.1. Comparative analysis

To evaluate the proposed scheme and the proposed descriptor, experiments were performed on two standard datasets: a) real world computing partnership (RWCP) sound scene dataset [21,12][1] and b) daily sound dataset [22]. The RWCP dataset contains environmental sounds that were recorded in an anechoic chamber through a microphone and DAT recorder at 48 kHz [21] and were later down-sampled to 16 kHz. The RWCP dataset contains a total of 9722 instances and 105 different non-speech dry source sounds. The dry source sounds are the sounds free from room acoustics [21]. In RWCP 90, the audio classes are grouped in 14 categories. Almost each class consists of 100 sound files in RAW format converted into WAV format. The daily sounds dataset contains all non-speech sound files in WAV format with a sampling frequency of 16 kHz [22], and the sound files present in the dataset were downloaded from the internet or recorded using a microphone [22]. The dataset comprises 1049 sound files grouped into 18 different sound classes. For fall detection experiments, we have recorded 100 fall sounds through human subjects with associated events i.e. scream, and/or object falling e.g. cup breaking etc. The fall events were recorded in a rectangular room measuring 7 m × 2 m with carpeted floor. We have particularly selected the night time for experiments, so that the recorded sounds have minimum amount of external interventions. The fall events were performed at a distance of 1–6 m from the microphones with different angles in order to reproduce realistically different fall patterns. The fall events comprise of falls on hands, sides, back, and knees. The fall sounds were then merged with the sound files present in the RWCP and daily sound repositories using the Audacity software[2] to mimic the real world environments. The isolated sound files were used for the classifier training purposes; whereas, the merged sound files were used for the evaluation purposes of the proposed framework.

To evaluate the fall detection performance, we compared our method against other state-of-the-art audio representation schemes, i.e., MFCC [14], acoustic-LBP [23], and LPC [24]. To evaluate the performance, quantitative measures including the precision, recall, F-1 score, accuracy, and error rates were used, and these measures are computed using true positive (TP), true negative (TN), false positive (FT) and false negative (FN) rates. For fall detection evaluation the classifier training occurs over the balanced training sets having 70% data for training and 30% data for testing purposes. For this we consider fall and panic sound examples as one class, and samples of various other environmental sounds present in the datasets as other class. The results presented in Fig. 3 shows that the proposed method has the highest recall, accuracy, and F1-score rates, as well as the lowest error rate as compared to the above mentioned schemes. However, for precision, acoustic-LTP and acoustic-LBP have a similar performance, and the foremost reason behind this performance overlap is that the precision measure ignores the false negative (FN) rates. However, by considering all other measures, we can conclude that the proposed method is far more efficient in terms of fall detection. Apart from the challenging nature of the audio streams in Daily sounds dataset, the results presented in Fig. 3 clearly shows the superiority of the proposed method in terms of the precision, recall, F1-score, and accuracy rates. The proposed method also has provided the lowest error rates that clearly signify the reliability. In addition, we have also compared our method against other state-of-the-art methods. Fig. 4, shows the comparison between our method and the methods presented in [12,25–31]. The results indicate that the proposed acoustic-LTP features are robust in terms of the audio stream representation and classification.

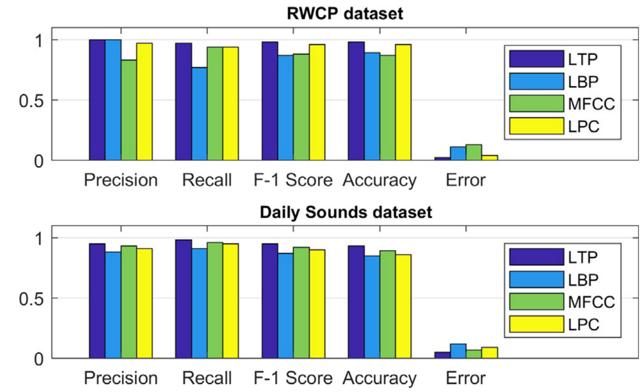

Fig. 3. Fall detection evaluation over RWCP and Daily sounds datasets.

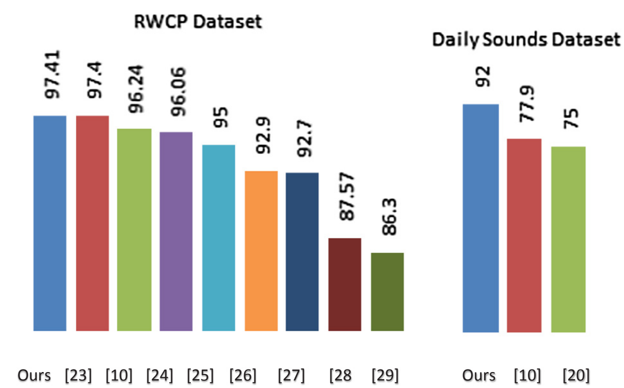

Fig. 4. Comparison in terms of accuracy over RWCP, and daily sounds datasets.

3.2. Evaluation in multiclass environment

To unveil the capabilities of the proposed acoustic-LTP as a reliable audio descriptor, we broadened the evaluation scope by performing the descriptors validation over the multi-category problems. In multiclass environment, the high correlation between the classes makes the instance recognition problem even more challenging. Due to the high correlation, the miss-association rate becomes high and class imbalance also raises a series of issues. In our work, the classification results over the multiclass problems are also computed and results are shown in Fig. 5. The proposed methodology achieves 92% accuracy on the daily sounds dataset and 97.41% accuracy on the RWCP dataset with the one-against-all classification settings for SVM classifier. Whereas, with similar experimental settings, MFCC achieves 88.6% accuracy on daily sounds dataset and 83.9% accuracy on RWCP dataset. Therefore, on the basis of evaluation results, it can be concluded that the proposed acoustic-LTP features, in multi-class problems, are robust in terms of audio stream representations as compared against state-of-the-arts.

3.3. Rotation invariance

For this experiment, the classifier is trained with audio samples having a rotation angle of 0°. Then we have used the audio samples

[1] Available online at: http://www.openslr.org/13/1.
[2] Available at the link: http://www.audacityteam.org/2.

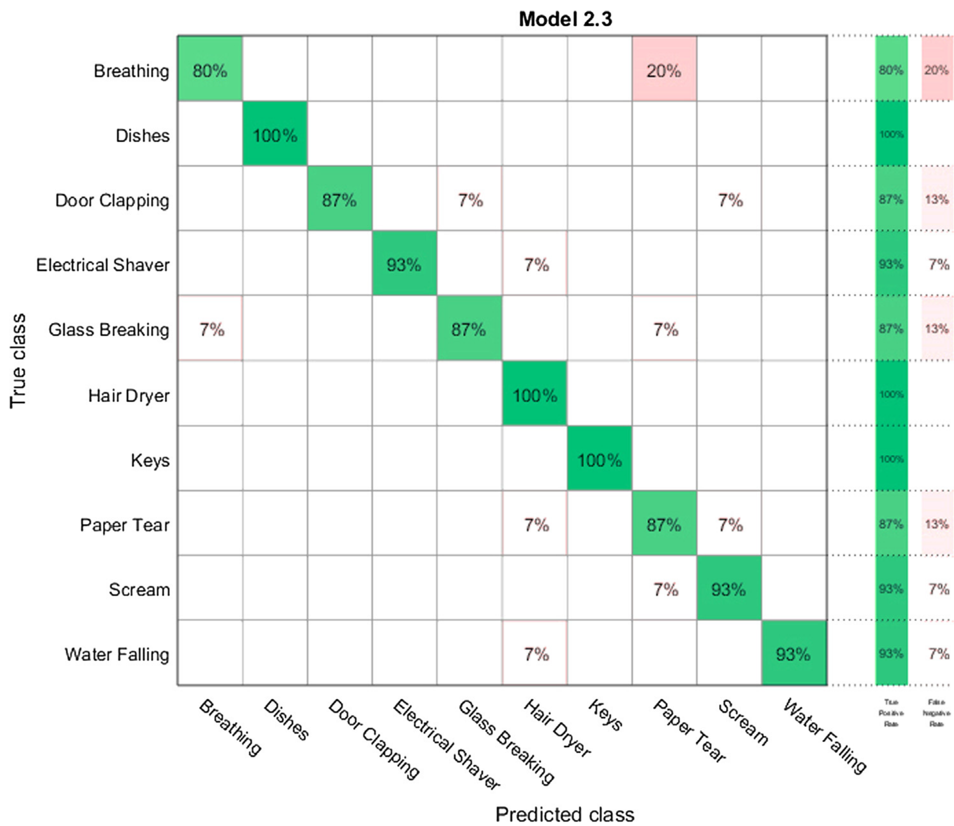

Fig. 5. Multiclass evaluation of acoustic-LTP over daily sounds dataset.

rotated 180° for the evaluation. It is important to note that rotation in audio signals allows representing semantically similar situations of expressions (i.e., expression of pain) in different tones.

Fig. 6 shows that the acoustic-LTP generates symmetrical representations for both the original and the inverted signals. To further investigate this important property, we randomly drew 30 fall and 20 non-fall samples and inverted the audio samples. The features for the inverted signals are extracted through the conventional and the proposed methods. Then, the classification for the fall and non-fall events is carried out through a trained classifier. The results of the classification are reported in Table 1. The results show that the proposed method outperforms other methods in terms of accuracy, F1-score, recall, and error rates. However, as precision measure ignores the FPs therefore, the acoustic-LBP, and MFCC perform better than acoustic-LTP. The results clearly reveal that proposed features are robust against rotated signals. The conventional approaches lack this feature. This could be

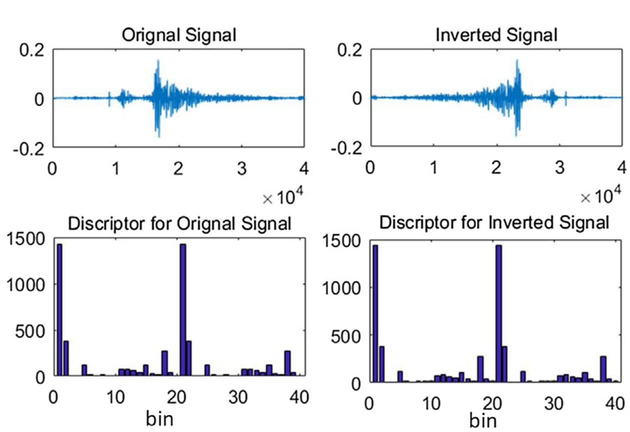

Fig. 6. Descriptors for original and inverted signals.

signified from the fact that for all these situations the audio representation through conventional descriptors will usually disagree. For example MFCC, which serves as almost a standard descriptor in audio applications generates different representations for high vs. low audio signals. Similarly if any event i.e. pain sound is expressed oppositely e.g. consider a scream that sounds “AAAH” against “HAAA”, the conventional approaches fail to recognize them as similar event and hence generate different feature representations that can also result in the form of misclassification. However, the proposed feature descriptor is able to address all these challenges in an effective manner. For evidence of the claim we have considered the screams that are expressed oppositely i.e. a scream that sounds “AAAH” (by breathing outside) is when rotated at 180° (or expressed oppositely e.g. by breathing inside) against “HAAA” (Fig. 6 first row), the audio representations remain constant (Fig. 6 second row); and hence, will be classified correctly. This constant representation for signals is a concept that is significantly focused in 2D signals domain i.e. in image representation in form of rotation invariance; whereas, it is an equally important concept in 1D signal domain particularly in critical applications. If a descriptor cannot address the rotation is prone to misclassified output and becomes a weak candidate for life impacting applications. Therefore, the proposed discriptor is more effective in fall detection.

4. Conclusion

In this paper, we have presented a framework for automatic fall detection for elderly people by analyzing the environmental sounds. Our fall detection framework has attributes of powerful audio feature extraction and representation mechanisms through acoustic-LTP. The proposed mechanism is more effective in classification and robust against the rotation attacks. A comparison of the performance against the state-of-the-art methods reveals the accuracy and reliability of the proposed method in terms of the fall detection to improve the quality of life for elderly people living an independent life.

Table 1
Classification Results over Rotated Signals.

	TN	FP	FN	TP	Accuracy %	Precision	F-1 score	Recall	Error rate
A-LTP	29	17	3	1	92	0.91	0.97	0.94	0.08
A-LBP	22	20	0	8	84	1.00	0.73	0.85	0.16
MFCC	13	20	0	17	66	1.00	0.43	0.60	0.34
LPC	19	9	11	1	56	0.63	0.95	0.76	0.24

Acknowledgement

This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2016R1D1A1B03933860).

References

[1] Stevens JA, et al. The costs of fatal and non-fatal falls among older adults. Injury prevention 2006;12(5):290–5.
[2] Gurley RJ, et al. Persons found in their homes helpless or dead. N Engl J Med 1996;334(26):1710–6.
[3] Doukas C, Maglogiannis I. Advanced patient or elder fall detection based on movement and sound data. IEEE; 2008.
[4] Grassi M, et al. A hardware-software framework for high-reliability people fall detection. 2008 IEEE Sensors. IEEE; 2008.
[5] Bianchi F, et al. Barometric pressure and triaxial accelerometry-based falls event detection. IEEE Trans Neural Syst Rehabil Eng 2010;18(6):619–27.
[6] Su BY, et al. Doppler radar fall activity detection using the wavelet transform. IEEE Trans Biomed Eng 2015;62(3):865–75.
[7] Liu L, et al. Automatic fall detection based on Doppler radar motion signature. IEEE; 2011.
[8] Wu M, et al. Fall detection based on sequential modeling of radar signal time-frequency features. IEEE; 2013.
[9] Ye J, Kobayashi T, Murakawa M. Urban sound event classification based on local and global features aggregation. Appl Acoust 2017;117:246–56.
[10] Glowacz A. Acoustic based fault diagnosis of three-phase induction motor. Appl Acoust 2018;137:82–9.
[11] Li Y, Ho K, Popescu M. A microphone array system for automatic fall detection. IEEE Trans Biomed Eng 2012;59(5):1291–301.
[12] Shaukat A, et al. Daily sound recognition for elderly people using ensemble methods. IEEE; 2014.
[13] Zigel Y, Litvak D, Gannot I. 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 2009;56(12):2858–67.
[14] Khan MS, et al. An unsupervised acoustic fall detection system using source separation for sound interference suppression. Signal Process 2015;110:199–210.
[15] Popescu M, Mahnot A. Acoustic fall detection using one-class classifiers. IEEE; 2009.
[16] Chen J, et al. Bathroom activity monitoring based on sound. Springer; 2005.
[17] Tan X, Triggs B. Enhanced local texture feature sets for face recognition under difficult lighting conditions. IEEE Trans Image Process 2010;19(6):1635–50.
[18] Bi Y, et al. Autodietary: a wearable acoustic sensor system for food intake recognition in daily life. IEEE Sens J 2016;16(3):806–16.
[19] Bingham E, Hyvärinen A. A fast fixed-point algorithm for independent component analysis of complex valued signals. Int J Neural Syst 2000;10(01):1–8.
[20] Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995;20(3):273–97.
[21] Nakamura S, et al. Design and collection of acoustic sound data for hands-free speech recognition and sound scene understanding. IEEE; 2002.
[22] Sehili MA, et al. Daily sound recognition using a combination of GMM and SVM for home automation. IEEE; 2012.
[23] Chatlani N, Soraghan JJ. Local binary patterns for 1-D signal processing. IEEE; 2010.
[24] Spratling M. A review of predictive coding algorithms. Brain Cogn 2017;112:92–7.
[25] Ozer I, Ozer Z, Findik O. Noise robust sound event classification with convolutional neural network. Neurocomputing 2018;272:505–12.
[26] Zhang J, et al. Robust sound event classification with bilinear multi-column ELM-AE and two-stage ensemble learning. EURASIP J Audio Speech Music Process 2017;2017(1):11.
[27] McLoughlin I, et al. Continuous robust sound event classification using time-frequency features and deep learning. PLoS One 2017;12(9):e0182309.
[28] Xie Z, et al. A new variance-based approach for discriminative feature extraction in machine hearing classification using spectrogram features. Digital Signal Process 2016;54:119–28.
[29] Guo X, et al. Environmental sound recognition using time-frequency intersection patterns. Appl Comput Intell Soft Comput 2012;2012:2.
[30] Peng Y-T, et al. Healthcare audio event classification using hidden markov models and hierarchical hidden markov models. IEEE; 2009.
[31] Mun S, et al. Deep neural network based learning and transferring mid-level audio features for acoustic scene classification. IEEE; 2017.