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
0003-682X/

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.

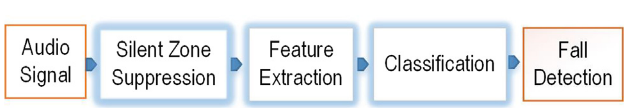

Fig. 1. Architecture of the proposed fall detection framework.

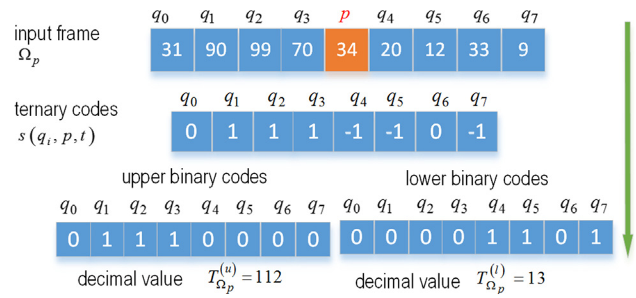

Fig. 2. Acoustic Local Ternary Pattern (acoustic-LTP) Computation.

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

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,

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