ACOUSTIC FALL DETECTION USING GAUSSIAN MIXTURE MODELS AND GMM SUPERVECTORS

Xiaodan Zhuang[1], *Jing Huang*[2], *Gerasimos Potamianos*[2], *Mark Hasegawa-Johnson*[1]

[1]Dept. of ECE, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA
[2]IBM T. J. Watson Research Center, Yorktown Heights, New York, USA

ABSTRACT

We present a fall classification and detection system to distinguish falls from other noise in the home environment using only a far-field microphone. We propose modeling each fall or noise segment using a GMM supervector, whose Euclidean distance measures the pairwise difference between audio segments. A Support Vector Machine built on a kernel between GMM supervectors is used to classify audio segments into falls and various types of noise. Experiments on the Netcarity fall dataset show that the proposed fall modeling and classification approach improves fall segment F-score to 67%, from 59% achieved by a standard GMM classifier. We also demonstrate that the proposed approach effectively improves fall detection accuracy by re-classifying confusable labels in the output of dynamic programming using the standard GMM classifier.

Index Terms— fall detection, Gaussian mixture model, GMM supervector, Support Vector Machine

1. INTRODUCTION

Assistance to dependent people, particularly elderly people, staying at an unsupervised environment has attracted increasing attention in aging societies [1]. One public health problem related to this group of people is the fall, which could impair their independence [2].

The challenges of fall classification and detection are at least three-fold. First, falls are inconsistent phenomena. In [3], Noury et. al. identified 10 kinds of falls that require attention and response. Second, falls happening in realistic environments are easily confusable with daily noise, such as dropping objects, moving chairs, closing doors and walking steps. Third, falls may overlap with background noise.

The research community has started to explore fall detection using different sensors [3], such as accelerometers,video cameras, microphones and their combination, mostly in relatively simple experimental set up. While different sensors excel in different applications, microphones have the merit of easy deployment and relatively low processing load. In particular, a far-field microphone avoids the obtrusiveness of some other sensors, such as accelerometers or close-talking microphones. However, no previous work that we know of uses only such acoustic signal in a realistic home acoustic environment.

Table 1. Sound classes for fall classification and detection

FA	sound resulting from the subject falling
ST	noise when the subject sits down on the chair, possibly leading to a bit of chair movement
CL	noise of clapping hands
GU	noise when the subject gets up from the floor
MP	noise of moving, putting, or catching an object
DO	noise of dropping an object on the floor
DN	noise of opening/closing doors
WK	noise of walking steps
MO	other noise, including speech and non-speech human voices, telephone ringing and other acoustically salient noise
BG	background noise, usually not perceptually salient

In this work, we focus on fall classification and detection using only acoustic signal. We use Gaussian Mixture Models (GMM) as our baseline. Motivated by progress in speaker identification [4], we propose using a Support Vector Machine (SVM) built on GMM supervectors to distinguish falls from other noise. In this approach, a universal background model (UBM) learns the shared acoustic feature space for falls and other noise, and the supervectors extracted from the GMMs adapted using each audio segment serve as robust summary of the acoustic signal. Experiments carried out on the Netcarity fall dataset [5], which carefully simulates falls and other activity noise in a realistic home environment, show that the SVM built on GMM supervectors improves both classification and detection of falls, compared with the baseline GMM approach.

The rest of the paper is organized as follows. Section 2 introduces the general framework for fall classification and detection. Section 3 discusses the GMM supervectors for audio segments. Section 4 derives the distance between the GMM supervectors and the corresponding GMM supervector kernel used in an SVM. Our experiments on the Netcarity fall dataset are presented in Section 5, followed by Section 6 as conclusion.

2. FALL CLASSIFICATION & DETECTION

Our fall detection systems identify existence and approximate occurrence time of falls. Segment boundaries of the acoustic input are found by the standard dynamic programming algo-

rithm. Each audio segment is classified into fall or various types of noise.

To better distinguish fall from all competing noise, we model falls and nine classes of noise in the living environment. These classes, shown in Table 1, are adopted with three considerations: Each class should have a sufficient number of instances in the training data. Each class is relatively distinguishable from others. The classes are chosen to better distinguish falls from noise.

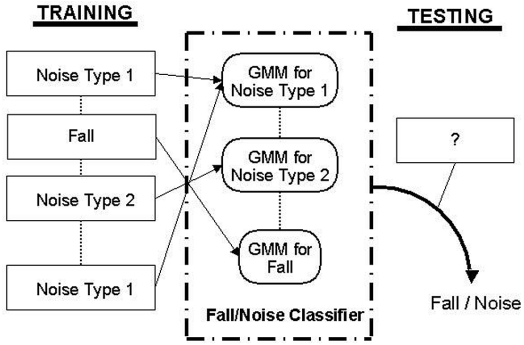

Fig. 1. Fall/noise modeling - the GMM approach.

Audio segment modeling is critical for distinguishing between fall and various types of noise. Each audio segment is represented by feature vectors extracted from evenly sampled and partially overlapping time-domain Hamming windows. Two approaches to model the distribution of these feature vectors are illustrated in Figure 1 and Figure 2. The first approach, or the standard GMM approach, approximates the joint distribution of all feature vectors in *each event class* with a GMM. For a test audio segment, a Maximum Likelihood classifier is used to obtain the hypothesized event class. We propose to use a second approach to model audio segments, referred to as the SVM-GMM-supervector approach, approximating the joint distribution of all feature vectors in *each audio segment* with a GMM, from which a GMM supervector is constructed as a summary of the segment. The pairwise Euclidean distances between these supervectors characterize the difference between the audio segments. Kernels derived from these distances are used in an SVM for classification.

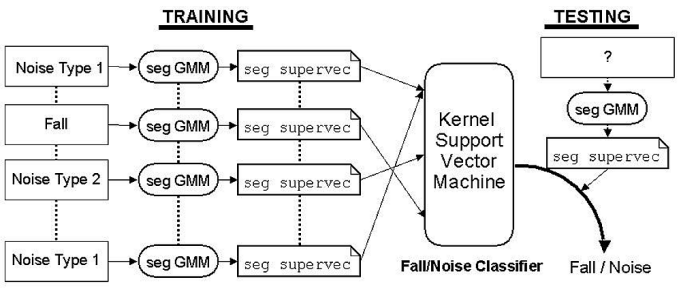

Fig. 2. Fall/noise modeling - the SVM-GMM-supervector approach.

3. GAUSSIAN MIXTURE MODELS AND GMM SUPERVECTORS

A Gaussian Mixture Model approximates the distribution of the observed features with a Gaussian mixture density function $g(z;\Theta) = \sum_{k=1}^{K} w_k \mathcal{N}(z;\mu_k,\Sigma_k)$, where w_k, μ_k and Σ_k are the weight, mean, and covariance matrix of the kth Gaussian component, and K is the total number of Gaussian components. The covariance matrices Σ_k are restricted to be diagonal for computational efficiency. The maximum likelihood parameters of a GMM can be obtained by using the Expectation-Maximization (EM) approach.

In Section 3.1 and 3.2, we discuss the construction of GMM supervectors.

3.1. UBM-MAP

Instead of separately estimating parameters for each GMM, we can also train GMMs by adapting from a global GMM, or a Universal Background Model (UBM). The potential merits of adapting GMMs from a UBM are two-fold. First, the parameters may be robustly estimated with a relatively small amount of training data. Second, there is correspondence between Gaussian components in different GMMs when these models are adapted from the same UBM.

More specifically, we obtain the GMMs by adapting the mean vectors of the global GMM using Maximum A Posteriori (MAP) criteria. The mixture weights and covariance matrices are retained for simplicity and robustness of parameter estimation.

MAP adaptation of GMM can be implemented by applying the EM algorithm. In the E-step, we compute $Pr(k|z_i)$, the posterior probability of the unimodal Gaussian component k given the feature vector z_i,

$$Pr(k|z_i) = \frac{w_k \mathcal{N}(z_i;\mu_k,\Sigma_k)}{\sum_{j=1}^{K} w_j \mathcal{N}(z_i;\mu_j,\Sigma_j)}, \quad (1)$$

where $w_k, \mu_k, \Sigma_k, k \in \{1,\ldots,K\}$ are the parameters of the UBM, and $Z = \{z_1,\ldots,z_H\}$ are the observed feature vectors. This step uses the UBM to assign each feature vector to the unimodal Gaussian components probabilistically, which establishes correspondence between the components of adapted GMMs, because the component parameters, i.e., the means, are estimated from statistics obtained involving the same UBM. In the M-step, the mean of each Gaussian component is updated,

$$E_k(Z) = \frac{1}{N_k} \sum_{i=1}^{H} Pr(k|z_i) z_i, \quad (2)$$

$$\hat{\mu}_k = \frac{N_k}{N_k+\tau} E_k(Z) + \frac{\tau}{N_k+\tau} \mu_k, \quad (3)$$

where τ is a weighting of the prior knowledge, i.e., the means in the UBM, to the observed data. In this work, τ is adjusted empirically according to the amount of available training data. N_k is the occupation likelihood of the observed data on the k^{th} Gaussian component: $N_k = \sum_{i=1}^{H} Pr(k|z_i)$.

3.2. Summarizing audio segments

Each audio segment is represented as an ensemble of feature vectors, extracted from $25ms$ Hamming windows with a step size of $10ms$. We calculate 12 Perceptual Linear Predictive (PLP) coefficients and the overall energy. On these 13 dimensions, utterance level cepstral mean subtraction is applied.

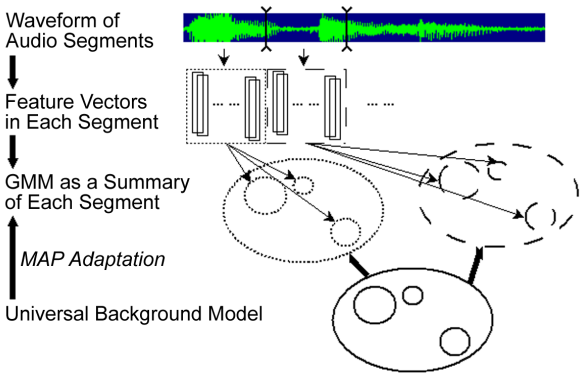

Fig. 3. GMMs (indicated by the ovals) summarize audio segments using multiple unimodal Guassians (indicated by the circles).

However, the feature vectors extracted from one audio segment carry a lot of noise. We use a GMM adapted from the UBM to capture the inner structure of the ensemble of feature vectors in each audio segment, as shown in Figure 3. According to Equation 1, the feature vectors are assigned to different unimodal Gaussian components probabilistically based on the UBM. We concatenate the adapted means of all the unimodal Gaussian components as a vector in a high dimensional space defined by the UBM, each dimension roughly corresponding to one dimension in the mean vector of one particular Gaussian component in the UBM.

This high-dimensional vector, called a GMM supervector, serves as a summary of the audio segment.

4. GMM SUPERVECTOR SPACE

4.1. Approximating Kullback-Leibler Divergence

As detailed in Section 3.2, we can summarize audio segments with supervectors constructed from GMMs adapted from the UBM respectively. We denote two such *segment* GMMs as g_a and g_b. A natural similarity measure between these two GMMs is the Kullback-Leibler divergence $D(g_a||g_b) = \int g_a(z) log\left(\frac{g_a(z)}{g_b(z)}\right) dz$.

The Kullback-Leibler divergence does not satisfy the conditions for a metric function, but there exists an upper bound using the log-sum inequality,

$$D(g_a||g_b) \leq \sum_{k=1}^{K} w_k D(\mathcal{N}(z;\mu_k^a,\Sigma_k)||\mathcal{N}(z;\mu_k^b,\Sigma_k)), \quad (4)$$

where μ_k^a and μ_k^b denote the adapted means of the kth component from the segment GMM g_a and g_b respectively. Since the covariance matrices are shared across all adapted GMMs and the UBM, the righthand side is equal to

$$d(a,b)^2 = \frac{1}{2}\sum_{k=1}^{K} w_k(\mu_k^a - \mu_k^b)^T\Sigma_k^{-1}(\mu_k^a - \mu_k^b). \quad (5)$$

We can consider $d(a,b)$ as the Euclidean distance between the normalized GMM supervectors in a high-dimensional feature space,

$$\phi(a) = [\sqrt{\frac{w_1}{2}}\Sigma_1^{-\frac{1}{2}}\mu_1^a; \cdots ; \sqrt{\frac{w_K}{2}}\Sigma_K^{-\frac{1}{2}}\mu_K^a], \quad (6)$$

$$d(a,b) = \|\phi(Z_a) - \phi(Z_b)\|_2. \quad (7)$$

4.2. Kernel for SVM

We use the GMM supervectors in an SVM for fall/noise classification. Since there are multiple types of noise, we tackle the problem as multi-class classification, implemented as binary classification problems via the one-vs-one method using LibSVM [6]. The distance defined in Equation 7 can be evaluated using kernel functions,

$$d(a,b) = \sqrt{K(a,a) - 2K(a,b) + K(b,b)}. \quad (8)$$

It is staightforward that the kernel function,

$$K(a,b) = \phi(a) \bullet \phi(b), \quad (9)$$

satisfies the Equation 8, where $\phi(a)$ and $\phi(b)$ are defined as in Equation 6.

5. EXPERIMENTS

5.1. Dataset

Our experiments are carried out on the acoustic fall data collected in the European project Netcarity [1, 5]. The dataset is of about 7 hours in 32 sessions, involving 13 different actors as subjects that might fall or perform other activities, and various other people that produce noise in the background. Figure 4 provides a snapshot. This dataset well simulates an environment that elderly people might encounter at home. We split the dataset into 20 training sessions, 7 testing sessions and 5 held out sessions for tuning the parameters. The subjects in the training and held out sessions do not overlap with those in testing. We map the labels in the Netcarity dataset to the ten classes detailed in Table 1 as the ground truth.

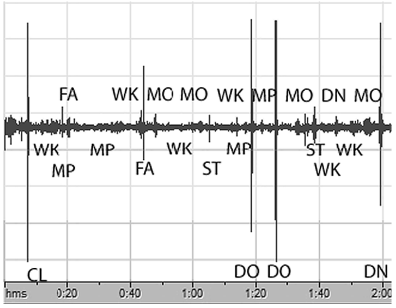

Fig. 4. Snapshot of Netcarity fall dataset (boundaries omitted for simplicity).

5.2. Experiment Setup

The first experiment is classification of audio segments whose boundaries are provided. Classification accuracy of all the ten classes defined in Section 2 reflects the overall performance of the classifiers. F-score of the fall segments reflects the capability to distinguish falls from all other noise. Both the GMM approach and the SVM-GMM-supervector approach are implemented with 512 Gaussian components for each GMM in this experiment.

The second experiment is detection of falls in acoustic signal of whole sessions. We measure the detection performance using AED-ACC [7], the harmonic mean between precision and recall, in which correctness is defined as the temporal center of either a hypothesized fall segment or a reference fall segment falling into the span of the other. In our experiment, we further require that all proposed fall segments not exceed a maximum length of 5 seconds so that the system output can be used for timely response to falls. Fall segments that exceed 5 seconds, if any, are removed from the output before scoring. We choose detection using the dynamic programming algorithm with the GMM audio segment modeling as our baseline. The SVM-GMM-supervector approach is adopted to re-classify the audio segments with perceptually confusable labels in the baseline output. In this dataset, the perceptually confusable labels are chosen to be falls (FA), dropping objects (DO), getting up (GU) and walking (WK).

5.3. Experiment results

Figure 5 illustrates the classification accuracy of all the ten fall/noise classes, and the F-score for fall segments. The results show that the SVM-GMM-supervector approach outperforms the GMM approach on classifying fall and noise segments.

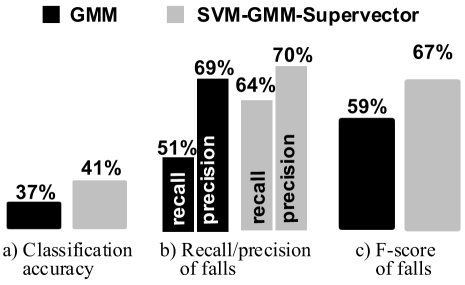

Fig. 5. Classification of falls/noise.

Figure 6 illustrates that using the SVM-GMM-supervector approach to re-classify confusable segments improves AED-ACC measure of the baseline output produced by dynamic programming with the GMM audio segment modeling.

6. CONCLUSION AND DISCUSSION

In this work, a fall classification and detection system is presented. We propose modeling each fall or noise segment using a GMM supervector and using an SVM built on a GMM supervector kernel to classify audio segments into falls and various types of noise. Experiments on a fall acoustic dataset collected from performed falls and other daily activities in a realistic home environment show that the proposed fall/noise modeling boosts classification performance, compared with a standard event class GMM classifier. The proposed approach also effectively improves falls detection accuracy by re-classifying confusable labels in the output of dynamic programming using the GMM classifier.

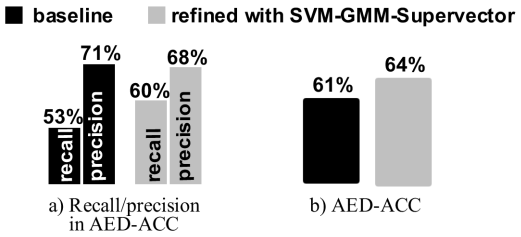

Fig. 6. Detection of falls.

Recent work in speaker verification applications [8] has shown further improvement using new classifiers based on GMM supervectors, compared to approaches similar to the SVM-GMM-supervector approach presented in this paper. This suggests possible further improvement in fall detection based on GMM supervectors.

7. ACKNOWLEDGEMENT

The authors would like to thank the authors of [5] for the Netcarity dataset, and Vit Libal and Larry Sansone for assistance with the dataset. This work is partially supported by the European Commission under integrated project NetCarity. The first author performed the work as a summer research intern at IBM T.J. Watson Research Center.

8. REFERENCES

[1] "Netcarity – ambient technology to support older people at home," http://www.netcarity.org.

[2] S. Sadigh, A. Reimers, R. Anderson, and L. Laflamme, "Falls and fall-related injuries among the elderly: a survey of residential-care facilities in a swedish municipality," *Journal of Community Health, 2004 Apr; 29(2):129-40*.

[3] N. Noury, A. Fleury, P. Rumeau, A.K. Bourke, G. Laighin, V. Rialle, and J.E. Lundy, "Fall detection - principles and methods," in *Proc: International Conference of the IEEE EMBS*, 2007.

[4] W. Campbell, D. E. Sturim, and D. A. Reynolds, "Support vector machines using GMM supervectors for speaker verification," *IEEE Signal Processing Letters*, vol. 13, no. 5, pp. 308–311, 2006.

[5] M. Grassi, A. Lombardi, G. Rescio, P. Malcovati, A. Leone, G. Diraco, C. Distante, P. Siciliano, M. Malfatti, L. Gonzo, V. Libal, J. Huang, and G. Potamianos, "A hardware-software framework for high-reliability people fall detection," in *Proc: IEEE Sensors 2008*, 2008.

[6] Chih-Chung Chang and Chih-Jen Lin, *LIBSVM: a library for support vector machines*, 2001, Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.

[7] Andrey Temko, "CLEAR 2007 AED evaluation plan," http://isl.ira.uka.de/clear07, 2007.

[8] Reda Dehak, Najim Dehak, Patrick Kenny, and Pierre Dumouchel, "Linear and non linear kernel GMM supervector machines for speaker verification," in *Proc: Interspeech 2007*, 2007, pp. 302–305.