Advanced Science and Technology Letters
Vol.145 (NGCIT 2017), pp.146-149
http://dx.doi.org/10.14257/astl.2017.145.29

Performance Comparison between GMM and SVM for Scream Sound Detection

Seok Hwan Jung[1] and Yong Joo Chung[1]

[1] Department of Electronics Engineering, Keimyung University, 1095 Dalgubul-Daero, Daegu, South Korea,
mester88@naver.com (Seok Hwan Jung), yjjung@kmu.ac.kr
(Yong Joo Chung: Corresponding author)

Abstract. In this study, we try to detect scream sounds based on SVM and GMM, respectively, and compare the performance of the two methods through various experiments. From the experimental results, SVM could obtain 0.559% False Acceptance Rate, which means that there is a very low possibility of incorrectly deciding non-scream sound as a scream sound. In contrary, the GMM method could achieve 12.03% of False Rejection Rate, which implies that GMM has relatively good sensitivity to the scream sound compared with SVM. From these results, we could conclude that both GMM and SVM have a distinctive merit from each other and the plausibility of further performance improvement by combining the two approaches is also observed.

Keywords: Scream detection, surveillance system, GMM, SVM.

1 Introduction

GMM(Gaussian Mixture Model) is one of the popular methods in audio detection as well as in speech recognition. GMM has shown successful results in detecting specific sounds such as scream, shout and gun fire [1][2][3]. Recently, researches using SVMs (Support Vector Machines) have been popularly used in audio detection [4][5][6]. SVMs are known to have equal or superior performance in generalization compared with other classifiers.

As mentioned previously, GMM and SVM are used independently in many researches to detect various audio signals but there are few research results to compare the two methods using the same audio data. But it is needless to say that we need to compare them to implement more efficient classifier for scream detection.

The paper is organized as follows. In section II, feature extraction methods and classifiers used for scream detection are introduced. In section III, we show and compare various experimental results. Finally, in section IV, we make conclusion and discuss further studies.

ISSN: 2287-1233 ASTL

2 Materials and methods

2.1 Feature Extraction

The waveform of sound signal can't be used as input of classifiers due to the irregularities in its characteristics. So some kind of values which can well explain the characteristics of the sound signal and traditionally, we use features like ZCR(zero crossing rate), pitch and correlation for audio signal detection [2][4]. In this study, we use MFCC features which have shown to be quite efficient with noise-robustness in speech recognition [1].

2.2 Gaussian Mixture Model

GMM is the sum of weighted Gaussian probability density functions and has been popularly used in speech recognition to model the acoustic characteristic of speech signals in MFCC domains. Training GMM in this study is done as shown in Fig. 2. After feature extraction, vector quantization is done for scream and non-scream data to find the GMM parameters for each of them. The GMM parameters consist of the weight ω_m , mean vector μ_m and covariance matrix Σ_m, $\{m = 1,2,\cdots,M\}$ of the Gaussian probability density functions comprising the GMM where M is the number of the mixture components.

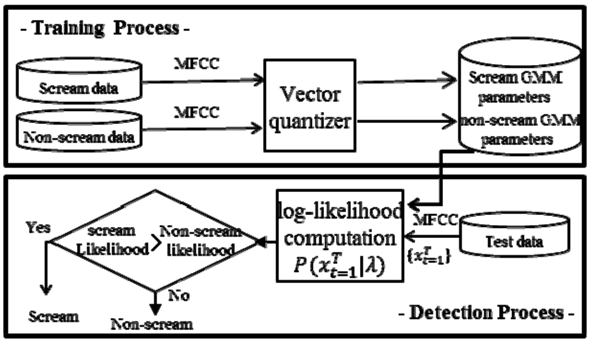

Fig. 1. GMM training and detection process.

2.3 Support Vector Machine

SVM is a non-probabilistic binary classifier which tries to maximize the distance margin between two classes [5]. In this study, the input for the SVM is obtained by averaging the 36 dimensional MFCC feature vectors for 20 frames.

3 Results

3.1 Databases

For the experiments in this study, we used data gathered from internet [7]. The data was recorded in clean conditions. The whole data can be divided into scream data set and non-scream data set. The scream data set consists of 63 files with durations from 1 seconds to 12 seconds. The non-scream data set consists of 213 files with durations from 1 seconds to 225 seconds.

3.2 Experimental Results

In Table 1, we show the results of GMM classifier when 12-dimensional MFCCs are used as feature vectors. We can see the results depend greatly on the mixture components of the GMM. FRR has the best result of 19.41% when the number of mixture components is 50. FAR improves as the number of mixture components increases and it has the best result of 9.1% when the number of mixture components is 76. The reason for the performance improvement of FAR with the number of mixture components is that the various sound signal in the non-scream data is modelled better as the number mixture components is increased. In contrary, the improvement of FRR with the number of mixture components is limited as the acoustic characteristic of the scream sound signal is relatively stationary.

Table 1. Experimental results using GMM classifier.

Mixture number	FRR(%)	FAR(%)
1	21.14	18.64
10	22.86	15.14
20	19.74	14.17
30	21.14	13.3
40	19.74	12.78
50	19.41	11.85
55	20.29	11.27
60	21.43	10.67
70	22.29	10.02
76	23.43	9.10

In Table 2, we show the comparison between GMM and SVM classifiers. For the GMM, the best results from Table 1 when the number of mixture components is 50(FRR) and 76(FAR) are shown in Table 2. From Table 2, we can see that SVM shows better performance in FAR(2.54% vs. 9.1%) while GMM is better than SVM in FRR(38.38% vs. 19.41%).

Table 2. Performance comparison between SVN and GMM.

	FRR(%)	FAR(%)
SVM	38.38	2.54
GMM	19.41	9.10

4 Conclusion

In this study, we compared the performance of SVM and GMM classifiers which are representative methods in audio detection. From the experiments, we could find that the two methods show contrary recognition results. GMM is superior in FRR while SVM is better than GMM in FAR.

References

1. S. G. Oh, J. U. Lee, H. S. Lee, Y. W. Chung and D. H. Park, "Abnormal Sound Detection and Identification in Surveillance System", Journal of KIISE, Vol. 39, No. 2, pp. 144-152, Feb. 2012.
2. J. H. Park, J. Y. Lim, J. Y. Yang, J. M. Kyung and M. S. Hahn, "False Positive Movie Clip Decision in Black-box Using Car Door-Closing Sound Classification", IEIE, Vol. 37, No. 1, pp. 761-763, June. 2014.
3. J. H. Seo, H. I. Lee and S. P. Lee, "A Design of a Scream Detecting Engine for Surveillance Systems", KIEE, Vol. 63, No. 11, pp. 1559-1563, Nov. 2014.
4. S. Ntalampiras, I. Potamitis, N. Fakotakis, "On acoustic surveillance of hazardous situations", Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on, pp. 165-168, April. 2009.
5. Support Vector Machines for Binary Classification, http://kr.mathworks.com/help/stats/support-vector-machines-for-binary-classification.html?requestedDomain=www.mathworks.com. [Accessed: Sep. 02, 2016]
6. Korean National Police Agency, http://www.police.go.kr/portal/main/contents.do?menuNo=200197 [Accessed: Sep. 05, 2016]
7. Sound Effects Download, http://www.soundsnap.com/ [Accessed: Sep. 05, 2016]