International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 4 Issue: 12 22 – 23

Intelligent CCTV Surveillance Based on Sound Recognition and Sound Localization

Min-Jeong Kim, Hyeon-Ji Yoo, Soo-Yeon Lee, and Seung Ho Choi*
Dept. of Electronic and IT Media Engineering
Seoul National University of Science and Technology
Seoul, Korea
e-mail: minjeongkim0411@hanmail.net, gmeing@hanmail.net, dlsy94@naver.com, shchoi@seoultech.ac.kr
**Corresponding author: Seung Ho Choi*

Abstract— CCTV is used for many purposes, especially for surveillance and fortraffic condition monitoring. This paper proposesan intelligent CCTV system that tracks sound events based on sound recognition and sound localization. From the experimental results, it is evident that the proposed method can be successfully used for the intelligent CCTV system of CCTV.

Keywords *- intelligent CCTV, surveillance, sound recognition, sound localization*

*****

I. INTRODUCTION

To eliminate blind spot, the need for intelligent CCTV has been increasing.This paper focuses on sond recognition and sound localization to make intelligent CCTV. To increase the accuracy of sound recognition, it is necessary to consider the appropriate noise-canceling technology. The spectral subtraction is commonly used for noise reduction in environments with various noises[1]. We use end-points detection algorithm to find the interval of the signal [2]. Then, mel-frequency cepstral coefficients (MFCCs) are calculated [3]. The the K-means algorithm is used for vector quantization (VQ) codebook[4, 5]. Finally, we use root mean square (RMS) to distinguish the left and right positions of the sound, and then find the correct angle and control the CCTV [6].

II. SPECTRAL SUBTRACTION

Spectral subtraction algorithms are widely used in speech enhancement. This method is to obtain the original signal by subtracting the spectrum of the noise estimated from the noise-added signal [1].However, when estimating a noise signal, a problem may arise in which an important part of the spectrumcan be regarded as noise. Therefore, to use the spectral subtraction technique, a sufficient interval is required to estimate the noise spectrum.Spectral subtraction results are shown in the figure 1.

III. ENDPOINTDETECTION

Endpoint detection is used to distinguish speech from noise and is required in many speech applications, such as speech recognition, speech coding and communication. In a speech recognition system, for example, accurate endpoint detection can improve the recognition accuracy under various types of background noise and reduce the computing power induced by incorrect speech detection [2].

IV. SOUND LOCALIZATION

Sound localization isto find where the sound is located[3]. For the sound localization, we used interaural time difference (ITDs), the time difference in which a signalarrives in two ears. In this experiment, two microphones are used to record sound. After selecting the location of the microphone, we used the time difference of the sound reaching the microphones,then calculate the angle of the direction of the sound and find out where the sound is coming from.

V. SOUND RECOGNITION

A. *Mel-frequency cepstral coefficients*

MFCC feature is commonly used for soundrecognition [4]. MFCC is one of the methods of expressing the power spectrum of the short-time signal. The frequency band of the MFCC can be evenly divided on the Mel-scale. So we can better express the sound.

B. *K-menas algorithm*

The K-means clustering algorithm isused for constructing VQ codebook [5, 6]. The algorithm divides the data set into several clusters. The K-means algorithm determines the sum of squares of the distance between the centroid of each cluster and updates the clusters in the direction of minimizing a cost function. We used the resulting VQ codebook for the sound recognition. Figure 2is an example of the codebook with 16 centroid vectors.

VI. EXPERIMENTS

We conducted the experiments with various sample sounds. Each sample has a mix of different types of noise, such as wind sounds and cell phone ring tones. The experimental results are shown in Table. 1.

TABLE I. SOUND RECOGNITION RESULTS (%)

sound type	**recognition rate (%)**
horn sound	92
alarm sound	83
collision sound	83
brake sound	75

VII. CONCLUSION

CCTV is widely employed for surveillance. However, it has blind spot because CCTV is fixed. In this paper, we proposed a sound event tracking algorithm based on sound recognition and sound localization to eliminate blind spot and increase the efficiency of CCTV. The experimental results showed that the proposed method can be successfully adopted for the intelligent CCTV system.

VIII. ACHNOWLEDGEMENTS

This study was supported by the Research Program funded by the Seoul National University of Science and Technology.

REFFENCES

[1] S. Boll, “Suppression of Acoustic Noise in Speech Using Spectral Subtraction,” IEE, vol. 27, pp. 113–120, 1979

[2] Wu, Bing-Fei, and Kun-Ching Wang. "Robust endpoint detection algorithm based on the adaptive band-partitioning spectral entropy in adverse environments." IEEE Transactions on Speech and Audio Processing 13.5 (2005): 762-775

[3] Jeffers, Lloyd A, “A place theory of sound localization,” Journal of Comparative and Physiological Psychology, vol 41(1), pp. 35-39, February 1948

[4] Zheng Fang, Zhang Guoliang, Song Zhanjiang, “Comparison of different implimentations of MFCC,” Journal of Computer Science and Technology, vol 16, pp. 582-589, November 2001

[5] Anil K. Jain, “Data clustering : 50 years beyond K-means,” Pattern Recognition Letters, vol 31, pp. 651-666, June 2010

[6] J. Wilpon, L. RAbiner, “A modified K-means clustering algorithm for use in isolated work recognition,” IEEE, vol 33, pp. 587-594, January 2003

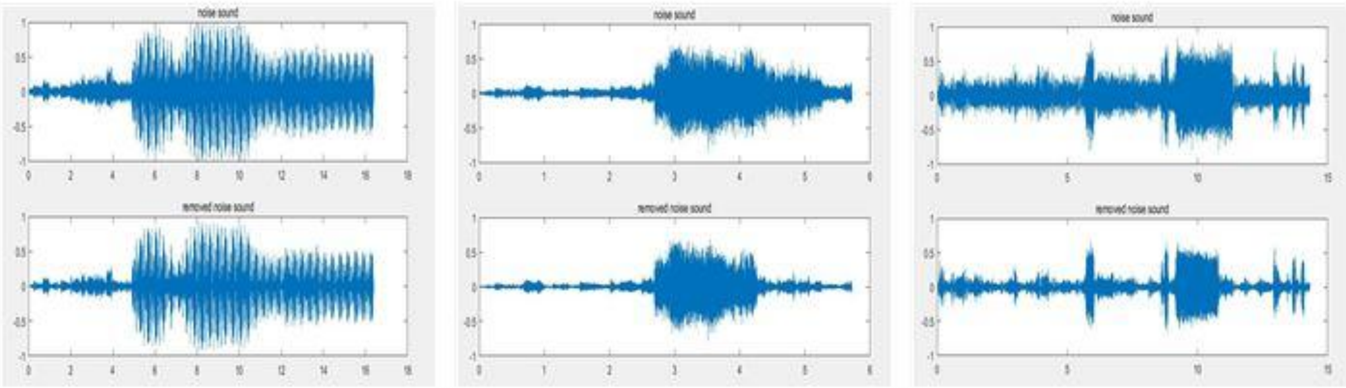

Figure 1. Spectral Subtraction Results: Noise-mixed alarm sound, Noise-mixed brake sound,Noise-mixed horn sound.

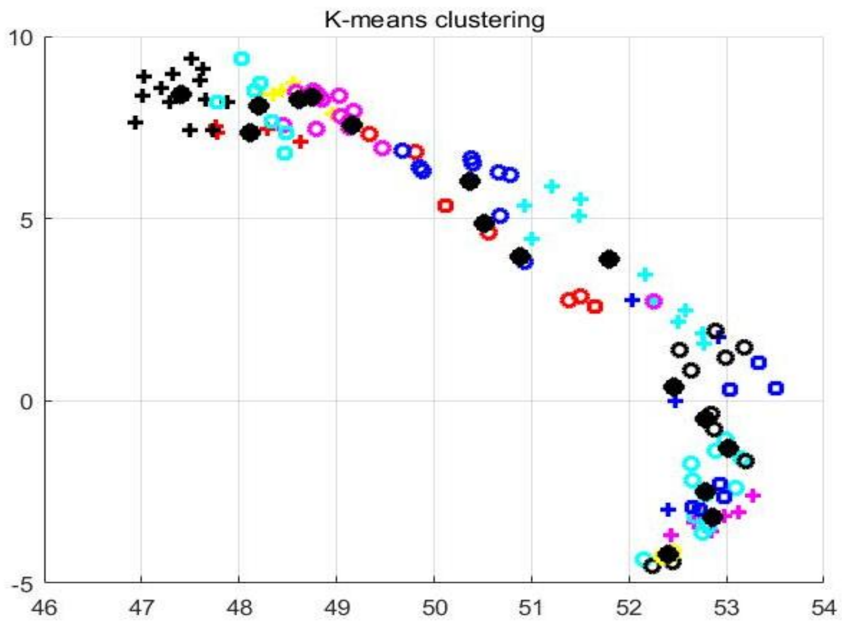

Figure 2. K-means clustering examples