A TV Commercial Monitoring System Using Audio Fingerprinting

Seungjae Lee and Jin S. Seo

Digital Contents Research Division, ETRI,
161 Gajeong-dong, Yuseong-gu, Daejon, Korea
{seungjlee,jsseo}@etri.re.kr

Abstract. In this paper, we describe an audio fingerprinting-based commercial monitoring system for TV broadcasting. The goal of the commercial monitoring is to identify commercials being broadcasted and find out the time duration. If there are similar commercials in the fingerprint database, we sometimes have trouble in determining which commercial has been broadcasted. To solve this problem, we propose the partial distance comparison method which identifies commercials and discriminates them from background music. Experimental results show that the proposed method improves commercial-identification performance by discriminating different versions of the same commercial.

1 Introduction

Audio fingerprinting is a technology which identifies an unknown audio with audio features, and it has a lot of applications including filtering for file-sharing services, automated monitoring for broadcasting stations, audio recognition through mobile network, etc [1]. In this paper, we proposed an audio fingerprinting-based commercial monitoring system which exactly identifies commercials when similar commercials are in the fingerprint database.

Similar commercials are produced by adding a new part, removing or replacing some parts of the original. Consequently, they have the same part, and we need to discriminate similar commercials. If all commercials have prior information such as title, we can classify them and use their information to identify the correct commercial. However, we can not get all information in advance, and there are sometimes errors in prior information. Therefore, we bind similar commercials by searching the fingerprint database when we create and update it. Based on those bindings, we can find candidate commercials and identify the broadcasted commercial by comparing partial distances.

2 System Overview

The proposed system is composed of recording server, searching server, and monitoring server as shown in Fig. 1. In recording server, the proposed system records broadcasting signal. In searching server, we search candidate commercials by using the extracted fingerprints from the recorded signal. In monitoring server, the system informs us the monitoring result among the candidate commercials.
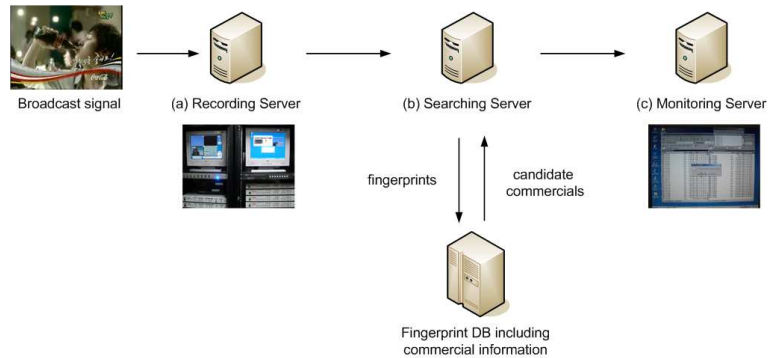
Fig. 1. The Commercial Monitoring System

2.1   Audio Feature

In commercial monitoring, the selection of audio feature determines the efficiency and the robustness of the system. Among previous work [2, 3], we use normalized spectral subband centorids[3] as audio feature because of the robustness against compression, equalization, random start, time-scale modification, and linear speed change. Instead of the original overlap ratio, we overlap 87.5% in order to get a fine time resolution. For each frame, 16 normalized spectral centroids are extracted, and 94 frames are used for the identification.

A set of commercials to be monitored is collected in advance, and then, the fingerprint database is generated based on the extracted features and their prior information such as title. At the same time, each commercial is searched in the present database in order to find similar commercials and bind them. We also make an index for the fingerprint database using K-D tree for a fast search.

2.2   Candidate Commercials

After extracting fingerprints from the recorded audio signal, we collect N sets of candidate positions in the database by searching the index tree using N successive fingerprints of the extracted fingerprints. Then, based on the sequential relationship among the N sets of candidate positions, the duplicated ones are removed and the unique set is obtained.

For the unique set, the Euclidean distance between the extracted fingerprints and the fingerprints in the database is calculated for 94 frames to collect candidate commercials. If there are commercials below the fixed threshold, we make a list and sort it by the time duration in descending order, and deliver it to the monitoring server. If not, we examine the next N successive fingerprints.

2.3   Decision and Time Verification

In monitoring server, we determine the title and the time duration of the broadcasted commercial by using partial distance comparison and time verification.

Let $D_{part}(p_i, q_i)$ be the partial distance at the ith position, and it is expressed as follows:

$$D_{part}(p_i, q_i) = \sum_{k=0}^{15}(p_i[k] - q_i[k])^2 \tag{1}$$

where $p_i[k]$, $q_i[k]$, and k denote the fingerprint of the candidate commercial in the database, the fingerprint of the broadcasting signal, and the frequency bin, respectively.

In partial distance comparison, we calculate the absolute difference D_{comp} between the left normalized sum of partial distance D_{left} and the right normalized sum of partial distance D_{right} is calculated as follows:

$$D_{comp} = |D_{left} - D_{right}| = \frac{1}{M - m}\Big|\sum_{i=m}^{M-1} D_{part}(p_i, q_i) - \sum_{i=M}^{2M-m} D_{part}(p_i, q_i)\Big| \tag{2}$$

where M and m denote the middle point of total frames to be compared and the number of frames to be skipped. We reject the first m frames and the last m frames to minimize an error caused by the boundaries between commercials. For the sorted candidate commercials, this comparison is continued until the D_{comp} is below the fixed threshold. If found, the time verification is performed. If not, new candidate commercials are collected by using the next N fingerprints.

In time verification, the start and the end position are determined by finding the minimum position of the sum of K partial distances. We use a 87.5% overlap ratio with 375.1ms frame, then the time resolution would be about 46.4ms.

3   Experimental Results

We made a fingerprint database from 1266 commercials. There are about 40% commercials which have similar commercials. The length of commercial is 5, 10, 15, 20, 30, or 60 sec. The test signal, which had been recorded from 2 different broadcasting stations, is about 12 hours, and it has 288 commercial appearances. Out of them, 269 commercials were in the fingerprint database at first, and 19 commercials were updated to it. We used 2.4GHz CPU and 2G memory. An average processing time for broadcasting signal of an hour is about 6 minutes. Table 1. shows the experimental results.

Table 1. The Commercial Monitoring Results

Test set	# of commercials	# of detection	# of false positive	# of false negative
1	288	269(21)	31	1
2	288	288	13	1

In the first experiment, without the partial distance comparison, the monitoring were performed by using the minimum Euclidean distance of candidate commercials. We exactly found 269 commercials, but there were 21 cases that needed the postprocessing to discriminate similar commercials. Moreover, some of false positive cases were occurred due to the audio similarity of a TV program and a commercial.

In the second experiment, we updated 19 commercials which are not in the database but in the test signal and tested the monitoring system with the partial comparison. As we expected, 288 commercials were detected without the postprocessing and false positive cases were decreased. However, there were still some false positive cases and a false negative case.

The false negative case had been occurred because of the overlap between the recorded broadcasting files. The false positive cases had resulted from the silence and noise-like sound of some commercials. To overcome this limitation, the monitoring system need some video features.

4   Conclusion

In this paper, we examined the commercial monitoring system for TV broadcasting. In commercials, there are a lot of similar commercials by adding a new part, removing or replacing some parts of the original commercial. As a result, we need to exactly discriminate each broadcasted commercial, and it is important due to business reasons. The proposed system introduces a commercial monitoring method based on the partial distance comparison whose discrimination performance was confirmed by experiments with real broadcasting signal.

5   Acknowledgement

References

1. J. Haitsma and T. Kalker. A highly robust audio fingerprinting system. In *Proceedings of International Symposium on Music Information Retrieval*, Ooctober 2002.
2. P. Cano, E. Battle, T. Kalker, and J. Haitsma. A review of algorithms for audio fingerprinting. In *Proceedings of International Workshop on Multimedia Signal Processing*, Decemeber, 2002.
3. J. S. Seo, M. Jin, S. Lee, D. Jang, S. Lee, and C. D. Yoo. Audio fingerprinting based on normalized spectral subband centorids. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, volume III, pages 213-216., March 2005.