Automatic Commercial Monitoring for TV Broadcasting Using Audio Fingerprinting

Dalwon Jang[1], Seungjae Lee[2], Jun Seok Lee[2], Minho Jin[1], Jin S. Seo[2], Sunil Lee[1] and Chang D. Yoo[1]

[1]*Korea Advanced Institute of Science and Technology, Daejeon, Korea*

[2]*Electronics and Telecommunications Research Institute, Daejeon, Korea*

Correspondence should be addressed to Dalwon Jang (dal1@kaist.ac.kr)

ABSTRACT

In this paper, an automatic commercial monitoring system using audio fingerprinting is proposed. The goal of the commercial monitoring system is to identify the title and the exact duration of commercials in real-time. To achieve this, only the audio is considered. The audio is easy to handle in real-time and can provide high accuracy for commercial identification. More precisely, the spectral subband centroids are extracted from an audio part of a commercial and indexed using the K-D tree algorithm. To detect aired commercials robustly, a four-step verification method using the indexed tree of the commercials is proposed. Experimental results show that the proposed system is robust against degradations during the real broadcasting and recording process and thus can fulfill the commercial monitoring satisfactorily.

1. INTRODUCTION

With the development of information and communication technology, we can access hundreds of thousands of multimedia files in the database (DB) , but searching through the DB for a file requires considerable computation. Various effective searching algorithms for multimedia files have been proposed in the literatures [1, 2, 3]. In audio fingerprinting [4, 5] audio feature like the human fingerprint is extracted and stored in a DB with meta data to identify the audio clip. It can be applicable to broadcast monitoring, music identification, and so on. Moreover, music identification services and broadcast monitoring services have been already deployed in some countries [6, 7].

In this paper, an automatic commercial monitoring system for TV broadcasting is proposed based on audio fingerprinting. In [8, 9], background music detection and monitoring is described using audio fingerprinting [9] and watermarking [8]. In this paper, the detection and monitoring TV commercials in real-time is considered. Commercial detection and monitoring seems to be similar to music identification, but its characteristics and requirements are quite different. First of all, commercials have very short duration between 10 seconds and 30 seconds. Thus, features from a short-duration of commercials are available for identification. Moreover, the monitoring system should report the exact duration and the frequency of an aired commercial to verify the contract between broadcasting companies and advertisers.

A commercial monitoring system should be robust against degradations which may occur during broadcasting or recording process. In this regard, the audio fingerprinting method [4] is applied to the proposed commercial monitoring system. The audio feature in the paper is known to be robust against various degradations. To calculate the duration time within 1 second error, the proposed system introduces a four-step searching process. Besides, the identical audio case is considered in proposed system. Since the identical audio may be used for different commercials, it is difficult for the system which has one-best output to detect a correct commercial. Though the proposed system can not solve the problem perfectly, it deals with this case by reporting list of candidate commercials. The proposed system binds all the commercials which have an identical audio in creating feature DB or updating feature DB and reports them altogether as a monitoring output. Experimental results show that the proposed system detects commercials with a high detection probability and a low false-alarm probability and reports all candidate commercials in similar audio cases.

For TV broadcasting, because video has more information than audio, the monitoring system using video feature is thought to be more proper than the system using audio. But, audio feature is chosen because processing audio data takes little computation than processing video data. Since there are the fundamental limits of audio, monitoring system using only audio features is not perfect. With the case of identical audio case mentioned above, the case that no information is extracted from sound is also a problem. The system can not detect the correct commercial if the commercial has no sound or noise-like sound. Though the cases rarely occur, those limit the system. To overcome the limits, it is assumed that the monitoring system using video must be made to complement the output of the proposed system. It is left as a further work.

This paper organized as follows. Section 2 presents general commercial characteristics and commercial monitoring system requirements. Section 3 describes audio fingerprinting feature to be used in the proposed system. Section 4 explains searching algorithm. Section 5 presents experimental results. Section 6 discusses conclusion and further works.

2. SYSTEM OVERVIEW

The proposed commercial monitoring system is presented in Fig. 1. It is composed of recording server, searching server and monitoring server, and each server performs the following functions:

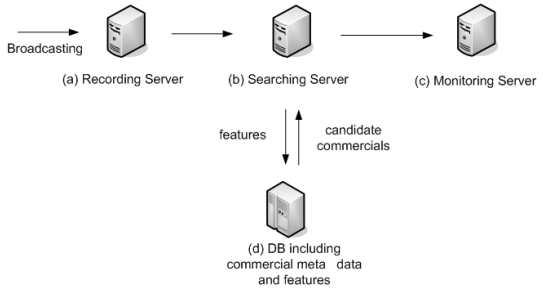

Fig. 1: The Block diagram of commercial monitoring system

- **Recording server**: To be tolerant of unexpected errors of the monitoring system, broadcasted audio data of a few hours should be recorded. After gathering audio data, a predefined amount of the recorded data is transferred to the searching server as an input. Therefore, the results of monitoring will be delayed by a few hours.

- **Searching server**: It can extract audio features from the recorded data and search the matched one from the commercial DB. It gathers the candidate feature positions from *DB search*, and chooses the closest commercial by *verification*. Moreover, the duration time is also calculated by *time verification*

- **Monitoring server** : It writes and updates the monitoring result sheets. The results include the title of the commercial, the starting time of the commercial, the ending time of the commercial, and the accuracy of the searching result.

For a liable commercial monitoring system, the characteristics of commercials should be considered. Usually, broadcast commercials have short duration and their content varies rapidly. A commercial can be classified into four different types as shown in Fig. 2.

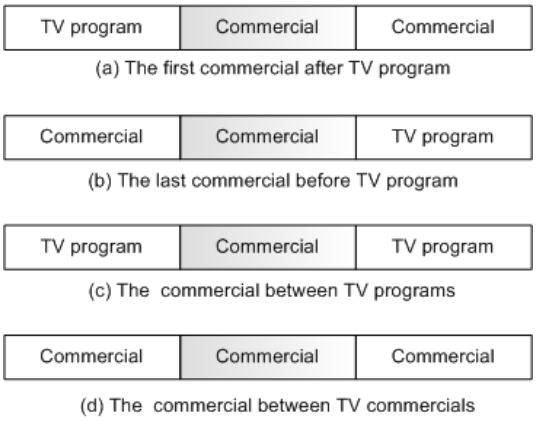

Fig. 2: Four different types of a commercial in broadcasting

The boundary search for each case is directly related to the starting time and the ending time, and the proposed system introduces time verification process and a longer overlapped frame for audio features in order to reduce the influence on the performance.

3. AUDIO FINGERPRINTING FEATURE

In audio fingerprinting, the selection of audio features which represent inherent characteristics of audio is a crucial problem because it can determine the efficiency and the robustness of audio fingerprinting system. A good audio feature should be robust against malicious attacks as well as attacks associated with signal processing while being separable from the features of the different audios.

Among previous works [4, 5, 10], spectral sub-band centroid is chosen as a feature for our system as in [4]. It has been shown that spectral sub-band centroid is robust against various signal processing such as MP3 compression, equalization, time-scale modification, and linear speed change that may happen in broadcasting. Moreover, it showed better performance than other well-known features such as MFCC and spectral flatness [10].

The frequency centroid of the 16 critical bands, which extracted from the downsampled audio signal frame, are used as a feature as in [4]. Instead of the original overlap ratio (50%) in [4], 75% overlap ratio is used for the more exact detection of time. This results in increasing the size of feature DB, but the precision of monitoring system is improved.

4. SEARCH ALGORITHM

After the audio feature is extracted, search process based on DB is performed. In the proposed system, the search algorithm consists of four processes: DB search, verification, decision, and time verification. The simple block diagram of commercial monitoring system is shown in Fig. 3. In the first three processes, the title of the commercial is identified. Then the starting and ending times of the commercial are determined in the fourth process. These four processes are explained in detail in the next subsections. In addition to these processes, the method for the commercials which have an identical audio is explained.

Basically, these processes are performed for every N_1 frames. But, once the title and the ending time of a commercial are detected, the frames in the corresponding commercial are skipped to reduce the processing time. The frames after the detected commercial are used in next search process. On the contrary, if the commercial is not detected, the frames after N_1 frames are used.

4.1. DB search

In DB search process, a set of candidates is selected using the tree-structure. For an effective search, K-D tree is used in the proposed method [1]. After searching the

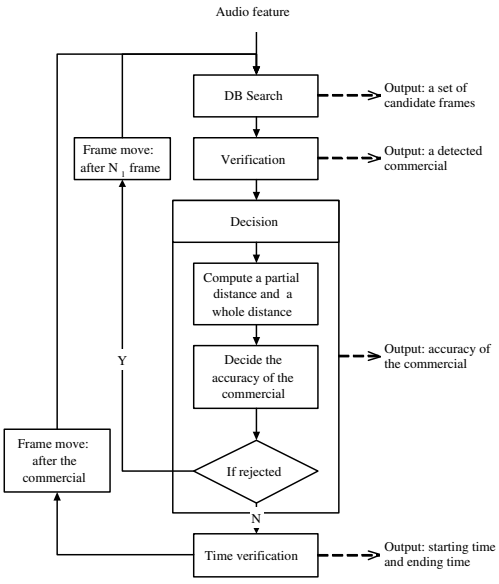

Fig. 3: Block diagram of searching algorithm

K-D tree for a frame, a set of candidate frames is obtained. The same DB search process is performed for $N_2 (N_2 \leq N_1)$ consecutive frames. From N_2 frames, N_2 sets of candidates are selected. But in the sets, there exist duplicate candidates. The word 'duplicate' in this case means that a candidate searched from a frame is p-th frame of a commercial and a candidate searched from the next frame is $(p+1)$th frame of the commercial. For a candidate, there can be at most $(N_2 - 1)$ duplicate candidate frames. Thus, the duplicate candidate must be eliminated while combining N_2 sets of candidates into a set of candidates. As N_2 gets larger, and N_1 gets smaller, the detection performance gets better, and the processing time gets longer. In proposed system, $N_1 = N_2 = 20$ is used.

In the tree, the feature of a frame, the title of commercial which the frame comes from, and the relative position of the frame in the commercial are stored. These information are used in the next processes.

4.2. Verification

In verification process, a candidate frame is chosen from a set of candidates using the Euclidean distance. From the candidate frame, the title of commercial can be determined. As written above, a frame is a basic unit for search in DB search process. But, in verification pro-

cess, a block which is composed of K_1 consecutive frame is used. The K_1 frames near the candidate frame are chosen in accordance with the relative position of candidate frame in the commercial. It means that the frames used in verification process should exist in a commercial. For example, if the relative position of candidate frame is the first part of the commercial, the frame and the next $(K_1 - 1)$ frames are chosen. If the relative position of candidate frame is the last part of the commercial, the frame and the previous $(K_1 - 1)$ frames are chosen.

After computing the Euclidean distance for each candidate, the candidate which has the minimum distance is chosen. If the distance for the candidate is smaller than the pre-fixed threshold, the candidate is determined as the final result of verification process. Even though the distance is larger than the threshold, the chosen candidate is verified once more in decision process. It will be explained in next subsection.

As the larger K_1 is used, the result is more accurate; however, since the commercial is broadcasted in a short time, the value of K_1 is determined in accordance with the length of commercial. In our work, the minimum length of commercial is assumed as 10 second. Accordingly, $K_1 = 104$ is chosen for the product of K_1 and frame interval, which is 92.875ms, (frame interval is 25% of frame length because of 75% overlap) to be shorter than 10 second.

4.3. Decision

In decision process, the verification result is verified once more. Decision process determines the accuracy of the verification result. Decision process uses the Euclidean distance as verification process does. By calculating partial and total Euclidean distance between query features and the detected commercial features, the accuracy is determined. To decide with partial distance, the distance for a frame and the distance for 10 frames are used. The total distance means that the distance of whole commercial between query data and the detected commercial. Even though the distance for K_1 frames is smaller than the threshold in verification process, the verification result can be rejected in decision process.

Decision process has four kinds of output as follows:

- **Accuracy level 1 (*perfect detection*)**: This level means the perfect detection of the commercial. If the both distances are within moderate values, the proposed system answers the result with high accuracy.
- **Accuracy level 2 (*suspicious 1*)** : In this case, the verification result is satisfied from the viewpoint of a total distance, but the partial distance has a peak value in a certain point. The commercial with this level can be considered as a new version of the existing commercial. There is a possibility that the recorded data may be degraded during broadcasting or recording time. This case should be verified using visual data.
- **Accuracy level 3 (*suspicious 2*)** : In this case, the total distance is out of a threshold, but the partial distance is within a threshold. That means a small part of the broadcasted data is similar to the commercial. Thus, there is a little possibility of being the commercial. The second verification using visual data is still necessary.
- **Rejection** : If the both distances are out of thresholds, the verification result if rejected.

If the rejection is determined, the search steps are repeated for next audio features as shown in Fig. 3. Of the four output, first three outputs present the accuracy of detected commercial. According to the thresholds and the number of frames used when computing partial distance, various decision levels can be made.

The candidate commercial which does not satisfy the threshold condition in verification process is also verified once more using only partial distance. In this case, the result is only verified as either suspicious 2 or rejection.

Decision result also helps to detect a new commercial similar to the existing commercial. New versions of some commercials are made by changing a part of audio of existing commercial. If the detected commercial is the slightly modified version of the existing commercial, it can be classified as a suspicious commercial in decision process. Those suspicious results help to understand a new commercial.

4.4. Time verification

As explained above, through the DB search process, not only the title of the candidate commercial, but also the relative position of present frame in the candidate commercial can be known. Thus, the starting time and the ending time of the commercial can be guessed with only result of previous process. But, sometimes, it is not trustworthy because the commercial can be cut from the original commercial or damaged by noise or some broadcasting signal. The companies who bought the airtime for

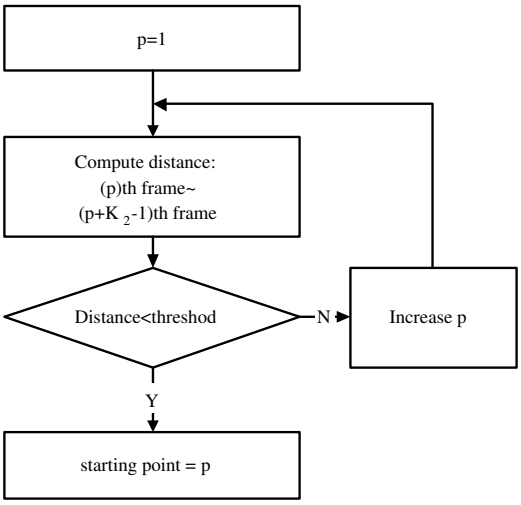

Fig. 4: Algorithm to find starting point

their commercials want to make sure that their commercials were broadcasted in time. Thus, one more verification process for the first part and last part of commercial is necessary. In this process, the partial distance for K_2 frames is used. Absolutely, K_2 is much smaller K_1. In our work, the value K_2 is chosen as 10 to verify about one second data. The algorithm to find starting point is shown in Fig. 4. The algorithm to find ending point is the reverse process of the algorithm of Fig. 4. The starting time or ending time is obtained after multiplying the starting point or ending point by frame interval.

4.5. Special case for identical audio

The output of the proposed system reports one best result. But there is a case in trouble if the output is the only one. The case is that visually different commercials have the identical audio content. To solve the problem, each commercial is bind to the other commercials which have the same audio of the commercial when creating or updating DB. For this, DB search process and verification process using total distance are performed when creating or updating DB. Through two processes, the commercials having same audio can be found and bind. Owing to binding, a set of commercials that have the same audio can be found directly when a best result is determined, and the list of commercials is reported. Among the commercials, the really broadcasted commercial can be detected using video feature, or manually if needed.

5. EXPERIMENTAL RESULT

There are 507 different commercials in DB. The length of clip is 10s, 15s, 20s, 30s, or 60s. There are some sets of commercials which are quite similar to each other. About 36 hours of real broadcasting data is used for test. The test data is gathered from 3 different broadcasting stations.

5.1. Processing time

Computer with 3.4GHz CPU is used in the experiment. An average processing time for broadcasting data of an hour is about 289 seconds. The broadcasting data which is input of the system is encoded by Window Media Video (WMV). The processing time includes the decoding time of WMV file.

5.2. Error probability

5.2.1. Probability of detection

The commercials stored in DB are broadcasted 459 times. The commercials that do not belong to DB are ignored. Our system finds all commercials without false detection of similar commercials. Among them, 446 commercials are decided as *perfect detection*. In other words, commercials of about 97.2% is detected perfectly. Five commercials are decided as *suspicious 1* and seven commercials are decided as *suspicious 2*. For these commercials, verification process using video feature is necessary. It is left as a further work.

If a similar commercial is in DB, the commercial out of DB can be detected as suspicious. For example, when 20-second commercial is broadcasted which is not in DB, the 15-second commercial which is an edited version of 20-second commercial is detected. This result is helpful when a new version of existing commercial which is slightly modified.

The detection of commercials which have an identical audio content was also successful. All commercials that have an identical audio were presented as an output.

5.2.2. Probability of false-alarm

Reporting detection of a commercial when this commercials were not aired occurred 19 times in 36-hour data. These commercials have silence or noise-like sound as audio data. In this case, it is not proper to use only audio feature. Thus, the special method to cope with silence or noise-like sound is necessary to make the monitoring system more trustworthy. This is left as a further work.

5.3. Accuracy of time

For all cases, the error for the starting and the ending time was under 1 second. It satisfies the goal of the proposed system.

6. CONCLUSION

The commercial monitoring system is constructed and tested by real broadcasting data. The system can not only detect which commercial is aired but also catch the starting time and the ending time of the commercial with high accuracy. Even though the commercial monitoring system using the audio fingerprinting detects the exact commercial with high accuracy, as written above, it can not be an independent system because sometimes the system can not make a trustworthy output. This is an essential limit of audio. For the complete system, the commercial monitoring system using video information is necessary. The construction of commercial monitoring system using video and the output of our system is left as a further work. The work to reduce false-alarm is also a further work.

7. ACKNOWLEDGMENTS

This work was supported by grant No. R01-2003-000-10829-0 from the Basic Research Program of the Korea Science and Engineering Foundation and by University IT Research Center Project.

8. REFERENCES

[1] C. Bohm, S. Berchtold, and D. Keim, "Searching in highdimensional spaces: Index structures for improving the performance of multimedia databases," *ACM Computing Surveys*, vol. 33, no. 3, pp. 322-373, 2001.

[2] J. Oostveen, T. Kalker, and J. Haitsma, "Feature Extraction and a Database Strategy for Video Fingerprinting," *Lecture Notes In Computer Science*, vol. 2314, pp.117-128, 2002

[3] A. Joly, C. Frelicot, and O. Buisson, "Feature statistical retrieval applied to content-based copy identification," in *Proc. Int. Conf. on Image Processing*, vol. 1, pp. 681-684, Oct. 2004.

[4] J. S. Seo, M. Jin, S. Lee, D. Jang, S. Lee, and C. D. Yoo, "Audio fingerprinting based on normalized spectral subband centroids ," *Proc. ICASSP 2005*, vol. 3, pp. 213-216, Mar., 2005

[5] J. Haitsma and T. Kalker, "A Highly Robust Audio Fingerprinting System", *Proc. ISMIR 2002*, Oct., 2002.

[6] Audible Magic Corp. ([Online]. Available: http://www.audiblemagic.com/)

[7] Shazam Entertainment Ltd. ([Online]. Available: http://www.shazam.com/music/portal/)

[8] T. Nakamura, R. Tachibana, and S. Kobayashi, "Automatic music monitoring and boundary detection for broadcast using audio watermarking," *Proc. Security and Watermarking of Multimedia Contents IV, SPIE*, vol. 4675, pp. 170-180, Jan. 2002.

[9] Y. Suga, N. Kosugi, and M. Morimoto, "Real-time Background Music Monitoring based on Content-based Retrieval," *ACM Multimedia 2004*, pp. 120-127, 2004.

[10] J. Herre, E. Allamanche, and O. Hellumth, "Robust matching of audio signals using spectral flatness features," *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 127-130, 2001.