ISSN: 2321-7782 (Online)
Special Issue, December 2013
International Journal of Advance Research in Computer Science and Management Studies
Research Paper
Available online at: www.ijarcsms.com

A Survey of Audio Fingerprinting

Punnami. MV[1]
PG Student
Department of CSE
Velammal Engineering College
Anna University
Chennai – India

Judith Sherin Tilsha. S[2]
Faculty
Department of CSE
Velammal Engineering College
Anna University
Chennai - India

Abstract: ***An audio fingerprint is a compact digest derived from perceptually relevant aspects of a recording. It is a unique identifier of an audio signal. There are popular audio fingerprinting schemes in a common framework with short query probes captured from microphone, which are surveyed and evaluated. The relevant areas to audio-fingerprinting include information retrieval, pattern matching, music cognition etc.***

Keywords: ***Fingerprinting, Thumbnail Generation, Hash Values, Fast Combinational Fingerprinting, Framing, Overlapping, Robustness.***

I. INTRODUCTION

Fingerprint systems are over one hundred years old; no two fingerprints will be same. Audio fingerprinting provides the ability to derive a compact representation which can be efficiently matched against other audio clips. Audio fingerprinting technologies have gained attention since the audio format is independently identified. A common use case is a query for example: in music recognition a user listens to a song in a restaurant, shopping mall or in a car, and wants to know more information about the song. The source of difficulty when identifying audio content derives from its high dimensionality and the significant variance of the audio data for perceptually similar contest. Instead of comparing the whole files the hash values should be compared.

II. PRIOR WORK AND MOTIVATION

State-of-the-art audio retrieval applications use a set of low level fingerprints extracted from the audio sample for retrieval. Haitsma et al. propose fingerprints based on Bark Frequency Cepstrum Coefficients. Highly overlapping frames are considered to ensure that the query probe can be detected at arbitrary time-alignment. Each fingerprint is 32 bits and can be compared efficiently with Hamming distances.

III. SURVEY FOR FINGERPRINTING SCHEMES

Before surveying popular audio fingerprinting schemes, we discuss the audio retrieval applications. First a set of fingerprints are extracted from the query song. Then the query is compared with the database of reference tracks to find candidate matches. To avoid pair-wise comparison between the query and all of the reference tracks, the database is partitioned. The partitioning of database is pre-computed with the list of songs. The partitioning of the database could be done by hashing of the fingerprints.

Papers Reviewed:

- Algorithms For Audio Fingerprinting
- Robust Audio fingerprint's based Identification Method

© 2013, IJARCSMS All Rights Reserved

- Robust Audio Fingerprinting System
- Duplicate Detection And Thumbnail Generation
- Shazam Algorithm for Audio Search

A. Algorithms for Audio Fingerprinting

The different techniques of mapping functional parts to blocks of a unified framework have been used here[1]. It focuses on identification.

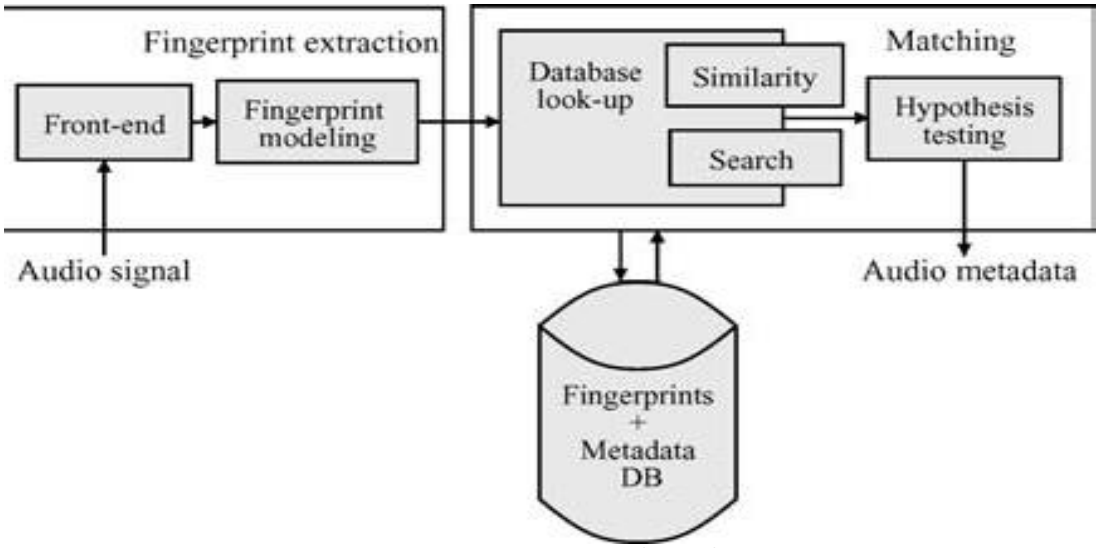

FIG1.1.content based audio identification framework [1]

The framework has two fundamental processes that are fingerprint extraction and matching algorithm. The extraction derives a relevant characteristic of a recording in robust and concise form. The extraction consists of front-end computes a set of measurements from the signal and a fingerprint modeling block defines the final fingerprint representation. The front end has five blocks.

a) Pre-processing - audio converted to raw format, to a certain sampling rate.

b) Framing &Overlap-signal can be stationary over an interval of milliseconds. The number of frames calculated per second is called frame rate. Overlap is applied when the input is not perfectly aligned.

c) Linear transform- set of measurements transformed to new set of features. Transformations are problem independent and computationally complex.

d) Feature Extraction-At the same time the dimensionality is reduced to increase the invariance to distortions

e) Post processing-It is to characterize the temporal variations in the signal.

Fingerprint Models-receives a sequence of vectors calculated on a frame by frame basis. Distances and Searching methods-Euclidean distance that deals with the sequence of different lengths, the issue in searching methods is the comparison of the unknown audio against the million fingerprints. Hypothesis testing- whether the query is present in the repository of items to identify.

B. Robust Audiofingerprint's based Identification Method

The robust features [2] are extracted and translated into a bit string; an object called a robust hash is obtained. The content can be identified by comparing hash values of a received audio clip with the hash values of stored original audio clips. A different feature of the proposed hash scheme is its ability to extract a bit string for every so many milliseconds. This schema is robust against severe compression, but bit errors will occur. Hash function is easily computable which maps the binary sequence

© 2013, IJARCSMS All Rights Reserved

of length M and binary sequence of length N; where (M, N) are the specified parameters. A hash function is used to summarize and verify large amount of data. The property of a hash function is its extreme fragility. Due to this property the cryptographic hash function is not suitable for summarizing multimedia, in which different quality versions yield the same hash value.

The two problems we identified here, firstly, a method is described for extracting a limited number of bits from the audio content. These hashing bits are robust, but not the probability of bits errors is close to zero. The problem here is the straight solution for non-exact pattern matching is an exhaustive search. Secondly, we describe a method that overcomes this exhaustive search complexity. By solving these two problems we trust that robust audio hashing is an efficient technology for audio identification.

Robust Audio Hash is a function that associates to basic unit time of audio content, a short semi unique bit sequence that is continuous with respect to content similarity as perceived by the Human Auditory System(HAS). It should be able to identify the content and the time interval. Two audio can differ drastically in a signal theoretic sense, whereas it is perceptually indistinguishable. There is no known algorithm to mimic HAS. The hash values for original content and for an MP3 compressed version of that content should be computed, the hash values should be similar. On the other hand, if two signals represent different content, the robust hash will be able to distinguish the two signals.

Robustness can be expressed by the Bit Error Rate (BER), which is defined as the number of error hash bits divided by the total number of hash bits. Instead of looking at all the 256 hash values in a hash block at once, we only look at one single hash value at a time and assume that every now and then that a single hash value has no bit errors. We start by creating a look up table for all possible 32-bit hash values, and letting the entries in the LUT point to the song(s) and the position(s) in that song where the respective hash value occurs. Since a hash value can occur at multiple positions in multiple songs the song pointers are stored in a linked list. Thus one hash value can generate multiple pointers to songs and positions. By exploiting the structure of an extracted hash block, the potential problem of an exhaustive search is reduced significantly.

C. Robust Audio Fingerprinting System

Audio fingerprinting [3] provides the ability to link short, unlabeled, snippets of audio content to corresponding data about that content. There are large number of applications for audio fingerprinting, ranging from identifying music based on cellphone playback, duplicate detection. Because of lossy compression of audio and many playback options, similar sounds may have largely different encodings, making simple matching is insufficient for this task. Different approaches have been attempted in the past, for approximate matching.

One of the most widely used systems use overlapping windows of audio, to extract interesting features. In overlapping, windows must be used to maintain time-shift invariance, for that exact time alignment is not known. 33 BFCC bands covering the 300-2000Hz range are used for the spectral representation. A sub-fingerprint is generated for every 11.6 milliseconds that cover a frame of 370ms. The large overlap ensures that the sub-fingerprints vary slowly over time.

The sub-fingerprints area vector of 32-bits indicate whether the difference in successive bands increases or decreases in consecutive frames. These sub-fingerprints are highly insensitive to small changes in the audio signal since no actual difference values are kept; instead, only the signs over continuous frames compose the sub-fingerprint. Comparisons of these fingerprints are efficient; they are simply the Hamming distance of the fingerprints.

D. Duplicate Detection and Thumbnail Generation

The two applications of audio fingerprinting [4]: duplicate detection, the goal is to identify duplicate audio clips in a set, even if one is a noisy version of other or if they have various durations. Duplicate detection is useful for automatically cleaning large audio collections. The second is thumbnail generation, whose goal is to provide a representative short clip of a music track i.e., the task is to find a short representative section of the music. It can help to improve audio browsing, either in simple

© 2013, IJARCSMS All Rights Reserved

plain list interfaces, or in more complex multidimensional ones. The two applications perform well, because of the robustness of the fingerprinting engine. In the previous work on thumbnails, a representative segment is searched for by maximizing similarity to all other segments in the clip.

In fixed length, segments are clustered, methods are used to choose the thumbnail, and results are given on 18 Beatles' songs. Both methods use Mel Cepstral features. Here, instead of using a feature set that has been trained to be robust against a variety of distortions, and the results on the overall quality of the thumbnail using blind testing on a larger data set. Two applications were built using the RARE (Robust Audio Recognition Engine) AFP system, which converts an audio segment to 64 floating-point numbers, and clips are identified using a weighted Euclidean distance. RARE has been shown to be very robust to distortions of the original audio. In the following, "trace" will mean any kind of fingerprint extracted from audio, and "fingerprint" is a reference fingerprint against which traces are compared to determine the audio recognizing.

The RARE duplicate detector recursively detects all audio files in a directory. For each file it creates a set of traces and checks them against a set of fingerprints from other audio files. If the normalized Euclidean distance between a fingerprint and a trace falls below a fixed threshold, the audio files associated are declared to be duplicates. The RARE audio thumbnail generator Gen Thumb whose goal is to find parts of the audio that repeats within the audio clip. If a song is chorus then all the instances of chorus are similar, then the system will be able to recognize the chorus, and use chorus to construct thumbnail. Gen Thumb uses audio fingerprints to find repeating sections, since expecting to generate fingerprints from similar sections of music. Rather than attempting to match the original audio using audio fingerprinting has two advantages:

a) Due to the robustness of RARE to distortions, variations of the same segment within a song will often still give similar fingerprints
b) Finger prints are low-dimensional representations of the original music ,so handling them instead of the audio is more efficient

E. Shazam Algorithm for Audio Search

This explores [5] the flexible audio search algorithm which is noise resistant, computationally efficient, capable of quickly identifying a short segment of music captured through a mobile device microphone in the presence of foreground voices and other dominant noise, and it is through voice codec compression, out of a database of over a million tracks.

The Shazam entertainment provided an algorithm with the ability to recognize a short audio sample of music that had been broadcasted, mixed with heavy ambient noise, subject to reverb and other processing captured by a cell phone microphone, subjected to voice codec compression, and network dropouts, all before arriving at our servers. The recognition over a large database of music with nearly 2M tracks should be performed quickly with the algorithm. Apart from music recognition over a mobile phone the Shazam algorithm is very fast and used in many applications. Due to the ability to dig deep into noise the music hidden behind a loud voiceover, such as in a radio advert can be identified.

Each audio file is fingerprinted in which reproducible hash tokens should be extracted. A large set of fingerprints derived from the music database are matched against the fingerprints generated from the unknown sample. The candidate matches should be evaluated continuously for correctness of match. To use the attributes of audio fingerprints some principles is available they should be robust temporally localized, translation-invariant, and sufficiently entropic. Robust Constellations - Registration of constellation is an efficient way of matching in the presence of noise or in deleted features.

Fast combinational fingerprinting-From the constellation maps the fingerprint hashes is formed in which a pair of time frequency band is combinatorilly formed. Anchor points are chosen and each point is associated with the target anchor point. Each anchor point is sequentially paired with points within its target zone. The hashes are reproducible, even in the presence of noise and voice codec compression. Furthermore, each hash can be packed into a 32-bit unsigned integer. Each hash is also

© 2013, IJARCSMS All Rights Reserved

associated with the time offset from the beginning of the respective file to its anchor point, though the absolute time is not a part of the hash itself.

Searching and Scoring: To perform search, audio fingerprinting should be done with the captured audio sample to generate a set of hash. Scoring is for matching hashes should be temporally aligned. Hash from the sample is used to search the matching hash from the database. For each matching hash corresponding offset values are calculated from the beginning of the sample and the database files are associated with time pairs. This algorithm performs well in the noise and even in the distortion. The speed of searching is within milliseconds.

IV. CONCLUSION

We perform thorough survey and evaluation of popular audio fingerprinting schemes such as searching, retrieval and its applications in a common framework. We discuss and report results important for Audio retrieval, size of fingerprints generated compared to size of the compressed audio sample and computational cost of fingerprint generation.

References

1. P. Cano, E. Batlle, T. Kalker, and J. Haitsma, "A Review of Algorithms for Audio Fingerprinting," The J. VLSI Signal Processing, vol. 41, no. 3, pp. 271-284, 2005.
2. J. Lebosse´ , L. Brun, and J.C. Pailles, "A Robust Audio Fingerprint's Based Identification Method," Proc. Third Iberian Conf. Pattern Recognition and Image Analysis, Part I (IbPRIA '07),pp. 185-192, 2007.
3. J. Haitsma and T. Kalker, "A Highly Robust Audio Fingerprinting System," Proc. Third Int'l Conf. Music Information Retrieval, Oct.2002.
4. C.Burges, D.Plastina, J. Platt, E. Renshaw, and H. Malvar, "Using Audio Fingerprinting for Duplicate Detection and Thumbnail Generation," Proc. IEEE Int'l Conf. Acoustics, Speech, and Signal Processing (ICASSP '05), vol. 3, pp. iii/9-iii12, Mar. 2005.
5. Wang, "An Industrial Strength Audio Search Algorithm," Proc.Int'l Conf. Music Information Retrieval (ISMIR), 2003.
6. Juels and M. Wattenberg, "A Fuzzy Commitment Scheme, "Proc. Sixth ACM Conf. Computer and Comm. Security, pp. 28-36,1999.
7. C.Bellettini and G. Mazzini, "A Framework for Robust Audio Fingerprinting," J. Comm.,vol.5,no.5, pp. 409-424, 2010.
8. Cheng Yang, "MACS: Music Audio Characteristic Sequence Indexing For Similarity Retrieval", in IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2001.
9. C.J.C. Burges, J.C. Platt, and S. Jana, "Distortion discriminant analysis for audio fingerprinting," IEEE Trans. on Speech and Audio Processing, vol. 11, no. 3, pp. 165–174, May 2003.
10. Erling Wold, Thom Blum, Douglas Keislar, James Wheaton, "Content-Based Classification, Search, and Retrieval of Audio", in IEEE Multimedia, Vol. 3, No.3: FALL 1996, pp. 27-36.
11. S. Sukittanon and E. Atlas, .Modulation frequency features for audio Fingerprinting,. International Conference on Acoustics, Speech and Digital Processing (ICASSP) IEEE, pp. pp II 1773.1776, 2002.

AUTHOR(S) PROFILE

Punnami MV, received the Bachelor degree in Information Technology from GGR Engineering College, Vellore in 2007-2011. Pursuing Master degree in Computer Science Engineering at Velammal Engineering College, Chennai. Works deliberately on the Audio Fingerprinting Survey for doing research on enhanced security methods.

© 2013, IJARCSMS All Rights Reserved