ISSN2394-3777 (Print)
ISSN2394-3785 (Online)
Available online atwww.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology
(IJARTET) Vol. 5, Special Issue 14, April 2018

AUDIO BASED SEARCH IN MYSQL CLUSTER ENVIRONMENT

DEVRAJ PATEL,
Department of Computer Science & Engineering National Institute of Technology Karnataka, Surathkal, Srinivasnagar,
Karnataka, India devraj.16cs06f@nitk.edu.in

VANI M, Associate Professor,
Department of Computer Science & Engineering
National Institute of Technology Karnataka, Surathkal, Srinivasnagar,
Karnataka, India vani@nitk.ac.in

ABSTRACT

Identification of audio using fingerprinting method have been an active research area since the fingerprinting techniques provides the detection of of audio without using its metadata or any other watermark support method. It is performed independently of its recorded format and depends on its embedded features. There are many other uses of audio fingerprinting techniques which includes integrity verification and content-based audio retrieval. Pattern matching, Music Information Retrieval or Robust Hashing are other terminologies being used for different approaches to audio fingerprinting. Few commercially deployed audio fingerprinting algorithms like Echoprint, Chromaprint and Landmark earmarks the significant progress already made in this area. However, there is room for improvement in the quest for more compact, robust and discriminate fingerprints and efficient searching algorithms. Enhancement of search speed and reduction of time space complexity are the challenges towards development of efficient and robust audio based search method.
With the advent of big data, there is a need to analyze high dimensional representations of the databases for faster processing of the matching content along with ensuring robustness against several audio distortions. It is proposed to create fingerprints using time frequency ratio with efficient hash function and implement it using multi-node clustering techniques. The clustering techniques will significantly reduce the number of hash comparisons and produces better time complexity of the searchalgorithm.
Keywords—Fingerprint, Metadata, Watermark, Content-Based Retrieval, Robust hashing, Landmark, Multi-node MySQL Cluster, Memory Efficient.

I. INTRODUCTION

There have been various advancements in communication technology towards incorporation of multimedia in all aspects of life. The amplifying need of managing huge data traffic also implies numerous copyright and duplication issues with the existing audio files. Content - Based Copy Detection (CBCD) is one of the concepts related to avoidance of copyright violation. It involves extraction of relevant audio features of the query file and then compares it against features of the original content stored in database.

Audio fingerprint technology is the generic term for the technique of transforming an audio signal based on its in-build features and create a reference database. When a query audio file is presented, the system identifies all the files similar to the query from the database (the stored signal, which are representations of the features extracted from original signals). There are various advantages of using fingerprints which includes reduction of the storage space and enhanced efficiency for identifying similar files[3].

The three well known fingerprinting algorithms are Echoprint, Chromaprint, Landmark. The Echoprint algorithm is an open-source audio fingerprinting algorithm, created and released by 'The Echo Nest' in 2011. It works by finding onsets points in time where audio notes occur. In Chromaprint algorithm, the audio file is firstconverted into time-frequency spectrum and then spectrum is converted to 12 bins, each representing the energy present in musical note. Audio hash is created by applying a set 16 filters. Once a query is given, all musical tracks that contains at-least one of the hashes in the database are retrieved. The Landmark algorithm extracts the peaks in the amplitude of the time-frequency spectrum. The audio track is divided into number of frames and then peaks are selected in pairs from each frame. Hashing is performed on the pairs formed by two frequency bins that lie between two points at time interval [14].
According to Alastair Porter [7], both Echoprint and Landmark produces similar results for small queries which

ISSN2394-3777 (Print)
ISSN2394-3785 (Online)
Available online atwww.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology
(IJARTET) Vol. 5, Special Issue 14, April 2018

starts from beginning of the stored tracks. However, Landmark gives better and consistent results with the randomly selected audio frame. The Chromaprint algorithm performs poorly with short queries(less than 8sec) but proved better for larger queries. In this paper, an efficient method is proposed for creation of audio fingerprints and its retrieval in multi-node mysql cluster environment. Section II of this paper, presents related work on audio fingerprinting. Section III describes the methodology used, Section IV explains the proposed plan of work and Section V gives experimental setup andresults.

II. RELATEDWORK

The fingerprinting algorithms defers in its types of features extracted from the audio files. In general, the fingerprinting method convert audio features to numerical values or hashes that becomes the value of a feature. Most of the fingerprinting algorithm computes the relevant features from an audio spectrogram using Fourier Transformation (FT) techniques. Using FT, frequency domain can be identified and represented from a time specific audio signal. In a thesis by A. Porter [7] a detailed comparative study and analysis has been mentioned with recommendations towards improvements in the field of audio signal fingerprinting and retrieval system.

In a paper by C. Ouali et. al. [11] the author mentioned another way to amplify fingerprint techniques by avoiding the search for irrelevant phases of the signal. The technique has been experimented and verified its application to the divide and locate (DAL) audio fingerprint method. In the literature, a signal search method called time-series active search (TAS) was proposed. TAS uses histograms of features, also termed as Bag of Features (BoF), as the fingerprint data. TAS has been used for broadcast monitoring for commercial messages. The author applied BoF over Divide and Locate (DAL) search method. The basic idea of DAL is to divide a spectrogram into a number of small regions and undertake matching for each region to locate it in the database. The small regions are quantized by vector quantization (VQ), and the matching operations are executed by comparing VQ codes.

According to paper on Fast Audio Fingerprinting System [12], author proposed an algorithm that identifies the transformed copy of an audio from a large collection of audio files in a database.. The audio fingerprints in this system converts audio signal into 2D binary images which encodes the positions of salient regions derived from its spectrogram. The intersection of these elements (i.e. positions of the salient regions) defines the similarity between two fingerprints. The search algorithm represents each reference fingerprint in the database with the closest query frame and then counts the number of matching frames when the query is mapped over the reference fingerprint. The best match is based on maximum count of matching frames.

According to Shazam Team [5], fingerprint hashes are formed using spectrogram (termed as constellation map) where time-frequency pairs are combinatorially associated. There are anchor points which are sequentially paired with points of its target zone. Each hash is also associated with the offset from the starting of the respective audio file. This process is carried out for each track in the database. In order to ensure faster processing, the structures are sorted according to their hash tokenvalue.

As per research of Petcharat et. al. [13], the time-frequency pair used by Shazam for creation of constellation map has been improvised by proposing a novel fingerprinting hash model using time-frequency ratio. The model is examined using a database composed of variety of eight genres of 300 audio clips, collected from ballroom dataset. The content based identification system (CBID) identifies matches based on musical content of the recordings in the corpus rather than the way that the file is stored by the computer. The author has proved better efficiency of the proposed model as compared to baseline model (as proposed by Shazam[5]). It was proposed to test the efficiency of the proposed model using other hashing techniques and different platform for database storage.

In a recent study by Milan Chikanbanjar [14], a detailed and step by step explanation of all the fingerprinting algorithm was carried out. The author explained the common steps involved in the audiofingerprinting systems. It includes preprocessing, framing & overlap, transform, feature extract, post processing and Fingerprint representation. The test was conducted using three fingerprinting algorithm (Echoprint, Chromaprint and Panako) on the test set of 300 recordings and observed that the precision of algorithm decreases as audio query is modified. The author emphasized that the audio fingerprinting is required for the detection of copyrights violation and finding duplicate audiofiles.

III. METHODOLOGY

A. Creation of fingerprints for all audiofiles

Landmark audio fingerprinting algorithm [14] is used for creation of fingerprints. Time-frequency spectrogram of each audio track is generated and peaks are extracted. Local maxima are extracted from a spectrogram of the

ISSN2394-3777 (Print)
ISSN2394-3785 (Online)
Available online atwww.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology
(IJARTET) Vol. 5, Special Issue 14, April 2018

audio frame. With the threshold created, prominent frequency peaks are determined. From each prominent peak, a set of spectrum feature pair is determined and combines to construct fingerprints. The number of audio fingerprints depends on the values of sampling frequency and degree of pairing with neighboring window while amplitude calculations. Each peak is paired with nearest neighbors in the declared fan out limit. The hash code has been generated by taking ratio of time and frequency from the spectrogram of the audio file. Important parameters are analyzed prior to fingerprinting process and their optimum values obtained are asfollows.

- Sampling frequency: According to Nyquist-Shannon Sampling Theory, 44100 samples per second is the theoretical limit on maximum frequency that can be detected for an audible stereo track. Hence, the considered value of sampling frequency in the paper is44100.
- Window Size for FFT: It is required to minimize spectral leakage while performing FFT. The size of window affects the frequency granularity, hence, depending on sampling frequency, 4096 is considered for this paper.
- Sequential window Overlap Ratio: It is the ratio by which each sequential window overlaps the last and next window. Higher value will represents more overlapping and leads to more number of fingerprints. Considered overlap ratio is0.5
- Peak Pairing : It represents the degree of pairing between peaks while computing hash. Larger value results in more number of fingerprints but also produces betteraccuracy.
- Minimum Threshold : It is the minimum number of amplitude required to be considered as peak. Higher number will reduce the number of fingerprints but affects the accuracy. Hence, optimal value considered for the experiment in this paper is10.
- Neighborhood Size : It is the number of cells around an amplitude peak in the spectrogram to be considered as spectral peak. Higher value reduces the count of local maxima hence leads to lesser fingerprints but it affects the accuracy of the algorithm. Optimum value obtained from the experiment is20.
- Hash reduction bits: It is the number of bits to be discarded from the front of the hash in order to reduce storage space. More bits leads to lesser storage but results in wrong identification ofsongs.

B. Store the fingerprints indatabase

The degree of pairing decides the number of sub fingerprints created for each track. Lesser value will reduce the space but compromise the search accuracy. There are tables in database for storing fingerprints of the processed audio files. Each pair of peaks gets fingerprinted using a hash function. Table contains three columns as Song ID, hash and Offset value. Each song track is converted into multiple frames and details of each frame isstoredindatabase.Theoffsetrepresentsthetimedifferenceofgiventrackfromitsoriginalsong.Each

C. Search matching audio fromdatabase

The query audio track is fingerprinted and compared with the hash values stored in database. All the similar hash codes along with song id and offset will be fetched thereafter the songid - offset pair which has maximum similarity with the query frame will be considered as best match of the particular query audio. While searching the query, the set of hashes is matched with the hashes stored in the reference database, and each exact match is returned. The matches are iterated while counting how many times each individual audio identifier occurs in the result set. Match with an audio identifier count lower than a certain threshold is removed and list of remaining audio identifiers is returned.

D. Configure Multi-node MySQLCluster

MySQL cluster provides auto-sharding and scalability with real-time responsiveness. This enables the database low latency with in-memory tables and indexes and binding of threads to CPUs. For the experimental analysis, MySQL cluster is configured in the lab with two client data node.

E. Performance Evaluation &Analysis

In order to evaluate the performance of a algorithm, the accuracy of the search results is observed. However if the reference database is big, i.e., if the number of songs fingerprinted and stored in the database are large, then time for matching is also considered. The main parameters are discussed here in detail with reference to search result.

- **Accuracy** : It is the ratio of correct match to the net total of queries that are presented for the same time duration. The audio queries are divided into five categories based of time in seconds. Each audio file is queried randomly from a directory and match time is recorded along withaccuracy.
- **Match Time** : It is the measure of time complexity which is required for fetching a result from the database. The match time varies depending upon the system configuration and methodology involved for database management. The total time taken to search the results isrecorded.

ISSN2394-3777 (Print)
ISSN2394-3785 (Online)
Available online atwww.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology
(IJARTET) Vol. 5, Special Issue 14, April 2018

IV. PROPOSED PLAN OFWORK

A. Algorithm for Fingerprintgeneration

The proposed audio search method uses the Landmark Algorithm [14] to convert songs into their respective hash codes. The morphological operations (iterate structure [1] and binary erosion [2]) used to find the local maxima and peak detection in the spectrogram. Each peak is paired with their neighboring peaks to form an anchor point for fixed number of peaks. Hash codes are generated from the time-frequency derivatives of consecutive peaks and theiroffset.

- FFT the audio channels and extract frequencycomponents.
- Generate 2D Spectrogram and apply log transform to normalize thearray.
- Replace infinite values in the spectrogram withzeros.
- Select minimum value of amplitude to be considered as peak and apply morphological operation to extractpeaks.
- ExtractpeaksfromlocalmaximainspectrogramafterXORingthelocalmaxwitherodedbackground.
- Consider the optimal number of peaks required to be paired for computation ofhash.
- Generate hash (sha1 or sha2) of peak frequency and time-frequency ratio of pairedpeaks.
- Discardfewinitialbitsfromhashsoastoreducethelengthoffingerprintsbeingstoredindatabase.

B. Algorithm for AudioSearch

The proposed audio search method uses the maximum similarity matrix for searching songs from the stored database. First, the query audio is fingerprinted to generate hash-time offset, then each hash is presented as query to search the matching hash from database.

- Create a dictionary of hash - offset pairs which will be used for fastersearch.
- Generate an iterator of all the reference hashes required to match the queryframes.
- Execute the query to fetch all the matching hashes withoffset.
- Create a list of hash-offset pair as matchresult.
- Once the list of matching hashes extracted from the database then each hash with time offset pair is aligned to get the exact searchresult.
- Exact result is found by counting the maximum occurrence of a particular id and offsetpair.
- Prepare a dictionary to collect the stored information of song id fetched asresult.

C. Size ofDatabase

The experiments were carried out on multi-node MySQL cluster environment [15]. Two data nodes and one SQL Server is configured as cluster. MySQL NDB Cluster is configured with NDB storage engine. NDB provides high availability and higher redundancy feature that are required for a database in distributed database environment. The dataset was collected from Ballroom dataset [16] and few online audio websites. The number of audio files taken and fingerprints created are as follows.

Table 1: Size of the database and details of fingerprints generated

Number of Songs	Audio Type	Length (sec)	Total fingerprints
5	Mp3	30	1,222,772
64	Mp3 , wav	180	4,473,730
664	Mp3, wav	>180	13,526,894
2000	Mp3, Wav	>180	35,018,548

V. EXPERIMENTAL SETUP ANDRESULTS

The algorithms are implemented using python and tested in lab using MySQL Multi-node cluster. Two data nodes and one master node are configured for the testing purpose. The test is performed to calculate the matching accuracy in percentage. Audio clips of varying length (1,2,3,4 & 5 sec) are extracted from the existing audio files in the directory and prepared a testing set. These audio files are then projected as query to the algorithm for identifying the matching files in the database. Bulk query presented in the single server takes comparatively longer time than the same when presented in cluster. The accuracy of respective query as per size is plotted.

ISSN2394-3777 (Print)
ISSN2394-3785 (Online)
Available online atwww.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology
(IJARTET) Vol. 5, Special Issue 14, April 2018

The accuracy test was conducted on randomly selected songs of type mp3 and wav from a directory. The result of matching accuracy are shown in Table II. The cluster provides higher potential to conduct concurrent processes which results in faster query search. Audio query is presented to single server as well as mysql cluster and remarkable reduction in match time is recorded for analysis.

Table 2: Accuracy of Searching algorithm on songs

Number of Query	Percentage Accuracy (as per query length)					Search Time in seconds
	1sec	2sec	3sec	4sec	5sec	
25	60	100	100	100	100	17.052
320	93.85	98.46	100	100	100	235.825
3320	88.08	99.26	99.41	99.85	100	1785.016
10000	86.4	97.55	98.8	99.35	99.75	8003.104

VI. CONCLUSION & FUTUREWORK

On experimental analysis, it is observed that the algorithm works efficiently on small database but performance decreases as the size of database increases. Identifying audio using fingerprint technique is much better and robust than using acoustic features like loudness, pitch or brightness. The audio fingerprinting technique with correlation of detected peaks as input to create hashes has resulted in reduction of number of reference hashes and hence reduces the storage space which further enhances the similarity matching time. Length of fingerprints (i.e. 32 bit or 512 bit) depends on the use of hash functions. However, the use of cryptographic hash algorithms do not affects the search accuracy as the value of hashes depends on time frequency spectrogram features.

In order to handle large database as reference fingerprints and to perform faster searching operation effectively with better accuracy, clustering technique is proposed as future scope. Multi-node MySQL Cluster would have been proved more effective provided the limitation of concurrent operations on the baseline systems be overcome. As the multiple query gets placed concurrently, it will get compared for its nearest cluster and then directed to the respective cluster to identify the exact match. The option of implementing KD Tree to improve search complexity can also be explored. This may reduce the time complexity and improve the system performance at great extent.

REFERENCES

[1] https://docs.scipy.org/doc/scipy/reference/generated/scipy.ndimage.iteratestructure.html
[2] https://docs.scipy.org/doc/scipy/reference/generated/scipy.ndimage.binaryerosion.html
[3] Pedro Cano, Eloi Batlle, Ton Kalker and Jaap Haitsma. ”A review of algorithms for audio fingerprinting” in International Workshop on Multimedia Signal Processing, US Virgin Islands, December2002.
[4] Erling Wold, Thom Blum, Douglas Keislar,and James Wheaton, “Content-Based Classification, search, and retrieval of audio” IEEE Multimedia, vol.3, 1996, pp.27-36
[5] Avery Li-Chun Wang, “An industrial strength audio search algorithm” Proceedings of the Fourth International Conference on Music Information Retrieval, Oct. 2003, pp.7-13.
[6] Ofir Lindenbaum, Shai Maskit, Ophir Kutiel and Gideon Nave, ”Musical features extraction for audio-based search” IEEE 26-th Convention of Electrical and Electronics Engineers in Israel,2010.
[7] Alastair Porter, ”Evaluating musical fingerprinting systems” Master’s Thesis, McGill University, Canada,2012.
[8] Parameswaran Vellachu and Dr. Sunitha Abburu, ”Tag based audio search engine” International Journal of Computer Science Issues, Vol. 9, Issue 2, No 3, March2012.
[9] Sunhyung Lee, Dongsuk Yook & Sukmoon Chang, ”An efficient audio fingerprint search algorithm for music retrieval” IEEE Transactions on Consumer Electronics, Vol. 59, No. 3, August2013.
[10] C. Ouali, P. Dumouchel, and V. Gupta, “A robust audio fingerprinting method for content-based copy detection” in Proc. Int. Workshop Content-Based Multimedia Index. (CBMI), Austria,2014.
[11] C. Ouali, P. Dumouchel, and V. Gupta, “GPU implementation of an audio fingerprints similarity search algorithm” in Proc. 13th Int. Workshop Content-Based Multimedia Index. (CBMI), Prague, Czech Republic,2015.
[12] Chahid Ouali, Pierre Dumouchel, and Vishwa Gupta, “Fast audio fingerprinting system using GPU and a clustering-based technique” IEEE/ACM Transactions on audio, speech, and Language processing, Vol. 24, no. 6, JUNE2016.
[13] Petcharat Panyapanuwat,Suwatchai Kamonsantiroj, and Luepol Pipanmaekaporn, “Time-frequency ratio hashing for content-based audio retrieval” IEEE2017.
[14] Milan Chikanbanjar, “Comparative analysis between audio fingerprinting algorithms” International Journal of Computer Science & Engineering Technology, May2017.
[15] https://www.digitalocean.com/community/tutorials/how-to-create-a-multi-node-mysql-cluster-on-ubuntu
[16] http://www.ismir.net/datasets.php

