International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-9 Issue-1S3, December 2019

Detection of Copy Right Infringement of Audio in on-Demand Systems using Audio Fingerprinting

Priscilla Rajadurai, M. Karthi, C. Niroshini Infantia, V. Neelambary, Lokeshwar Kumar Tabjula

Abstract: Today, with an evolution of large public media databases, repositories and services such as Sound Cloud, Spotify, Wynk, YouTube and Saavn, the chances of intentional and unintentional plagiarism is been increasing at an exponential rate. The paper proposed a solution which is used to detect copy right infringement of audio in real time systems using the technique called Audio Fingerprinting. The resulting software will be able to process an audio file, generate an audio fingerprint that can be stored and searched to find the matched values in a database.

Keywords : copy right, audio fingerprint, digital signal processing, music, hash function, plagiarism detection, license, audio database.

I. INTRODUCTION

When audio or video plagiarism cases are brought to court, independent media experts and analysts are asked to analyze the similarities and differences between the copyrighted song and the uploaded song, and their opinion plays a major role in detecting the plagiarism. In order to make their work easy, this paper gives a software solution with efficient techniques which will identify whether the uploaded song is copy righted or not. A lot of research is done so that the users meet the modern requirements. Now that users can interact with computers directly through voice, audio recognition technology has become a very competitive technology[1].

Let us consider a scenario where a high expectation official song or a trailer is going to be released soon in any one of the popular social media platforms like YouTube. People who perform copyright infringement may try to upload the same during the time of release, which will lead to monetization issues and ownership conflicts. This can be solved if the official content is copy righted and the duplicates are detected at the instant when they are uploaded.

Any three-second segment of the song is to enough for the system to detect whether the uploaded content is copyrighted or not. This is possible with the process of Audio fingerprinting. Audio fingerprinting is the process of extracting and identifying unique features of an audio content. It is a condensed digital data which is used to identify an audio sample or to place similar items in an audio database quickly that are generated from an audio signal**.** This content based identification system [2] can be used in copyright compliance and licensing schemes.

II. REVIEW OF RESEARCH

The design principles for the copyright infringement system are recurrent in recent research works. Fast indexing and retrieval is processed by employing Compact signatures in Information Retrieval that represents complex multimedia objects. The dimensionality should be reduced for performing indexing and searching at reduced space in order to handle complex multimedia objects [3]-[5].

One of the methods of audio fingerprinting include Multiple Hashing(MLH) with Discrete Cos transform to calculate the energy deference between the frames of the audio signal. Even though the quality of the segment is degraded, the recognition rate of the audio signal is improved. The size and the storage of the fingerprint are inversely proportional. But this increases the search time, hence the method was not adopted[6].

There are few systems that involves computer vision techniques like the wave print identification system. This system is based on wavelets and uses less amount of memory and hence can be used for extraction of fingerprints. Through this method, multiprocessing techniques are applicable which in turn reduces the time required to extract the fingerprints[9].

To extract the robust features from the audio segment, PRH is used which is most efficient approaches of audio is fingerprinting. The comparison of these hash values can be used to identify the unknown audio content. In this proposed system, the time difference between every frame of the audio segment is taken for hashing. That gives a 32-bit sub-fingerprint for each and every windowed signal[10].

Manuscript published on 30 December 2019.
* Correspondence Author (s)

***Dr. Priscilla Rajadurai,** Department of Information Technology, St.Joseph's Institute of Technology, Chennai, India. Email:prisci.christa@gmail.com
M. Karthi, Department of Information Technology, St.Joseph's Institute of Technology, Chennai, India. Email:mmuthukarthi@gmail.com
C.Niroshini Infantia, Department of Information Technology, St.Joseph's Institute of Technology, Chennai, India. Email: niroshini.siddarth@gmail.com
V. Neelambary, Department of Information Technology, St.Joseph's Institute of Technology, Chennai, India. Email:neelambaryvasu@gmail.com
Lokeshwar Kumar Tabjula, Department of Information Technology, St.Joseph's Institute of Technology, Chennai, India. Email: lokeshwartabjula@gmail.com

Retrieval Number: A10821291S319/2019©BEIESP
DOI:10.35940/ijeat.A1082.1291S319
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)
© Copyright: All rights reserved.

III. AUDIO FINGERPRINTING

A fingerprint is an output of a hash function that maps to an audio object. The audio object can be uniquely identified from a bit string. Hence basically an audio fingerprint comprises of a huge number of audio objects which are converted into bit strings by passing them into a hash function. The output of these hash functions are stored as bit strings into the database.

Hash functions compare the hash values H(X) and H(Y) where X and Y are large objects. For a given fingerprint function F, there occurs a threshold T such that (i) With high probability, $|F(X) – F(Y)|<T$ where X & Y are similar otherwise (ii) $|F(X) – F(Y)| > T$ if both the variable X & Y are dissimilar.

IV. SYSTEM DESIGN FLOW

Audio Copyright infringement system contains four major processes, where feature extraction plays a significant role. The efficiency of the system in terms of accuracy depends on the type of feature that is extracted from each audio object. The efficiency in terms of speed depends on the storage and retrieval techniques used in the system. They are discussed in detail further

A. Fingerprint Extraction

If the input is in the form of an audio signal, it is directly sent for preprocessing, else if it is a video, the audio alone is extracted and then preprocessing is done. A lengthy sequence of numbers is obtained when the music is digitally encoded. Each channel in an uncompressed .wav file contains 44100 samples per second and as a result a song of 3 minutes contains almost 16 million samples.

Multi-track recording and sound reinforcement are the operations performed by the audio signal communication channel known as audio channel or audio rack. To keep it simple with respect to this context, a channel serves as a medium that a speaker can play with a separate sequence of samples. Two channel setups are called as “stereo” and single channel setups are called as “mono”.

Now a days, many more emerging technologies yields a support to channels that are concurrently used in modern systems. The proposed system supports two channel setups. The 44100 samples per second is derived as per Nyquist-Shannon Sampling Theorem. The Nyquist Sampling Theorem states that: A bandlimited continuous-time signal can be sampled and perfectly reconstructed from its samples if the waveform is sampled over twice as fast as it’s highest frequency component. In the case of recording audio, frequencies above 22050 Hz is not necessary as humans are not possible to hear above 20,000 Hz frequencies. By applying the Nyquist-Shannon Sampling theorem, highest frequency should be sampled two times (ie) Samples per sec needed which is equal to Highest Frequency multiple of 2. The audio file is converted into digital signals and sub-frame where there lies an overlapping rate among the frames. This process is known to be audio file codec processing. Each frame contains multiple “fingerprint” segments Fast Fourier Transformation is used recursively on the song’s samples over a small intervals of time to create a spectrogram.

The frequency & time values are discrete that represents a bin whereas the amplitude values are real in nature. The color represents the real value of the amplitude at the discretized coordinate. In the spectrogram shown above, the red color represents higher values and green represents lower values.

B. Feature Extraction:

The spectrogram represents amplitude as the function of series of frequencies as the signal varies with time that signifies as two dimensional array. The column in Fast Fourier Transform shows the signal’s strength i.e., the amplitude at the particular frequency. The system does this enough number of times with a sliding window of FFT and puts the results together as a two dimensional spectrogram.

The robust way of capturing unique fingerprints from an audio signal is through “Peak Finding”. The greatest time and frequency pair associated with an amplitude value in a local neighborhood is termed to be a peak. The other pairs with lower amplitude are less likely to persist noise. Peaks are extracted through a combination of a high pass filter which accentuates high amplitudes and few local maxima structs from SciPy which an image is processing toolkit written in Python programming language. A unique hash is created by combining the peaks for a particular moment using their discrete time and frequency in the song.

A hash representing a unique fingerprint can be created with the combination of spectrogram peaks along with their time difference.

C. Store Fingerprint

The method that the system uses to store the generated fingerprints plays a significant role in the performance of the system in terms of speed. This is achieved by creating hashes of the fingerprints. The input and output of a hash function is an integer. A good hash function returns same output with the same or very few different inputs. In this system, the information of the pair of peaks and the time delta between them are given as input to the hash function as shown below:

Hash_Function (frequencies of each peaks, time difference between peaks) = Fingerprint Hash Value (Binary)

The algorithm used to create the hash function is SHA-128. By taking more than a single peak’s details into account, the system is able to create fingerprints with more entropy and henceforth contains more information. Thus they are powerful identifiers of songs since they will collide less. On one hand, more peaks in a fingerprint would easily identify a song but it also leads to have less robust in the face of noise.

The fingerprints table in the proposed system’s database schema will have the fields like Hash, song_id,offset INDEX(hash), UNIQUE(song_id, offset, hash). The table has a hash, a song ID and an offset that is associated with the time window from the spectrogram where the hash originated from. This will be required during the filtering process of matched hashes. Only the hashes that align will be from the true signal that needs to be identified.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)
© Copyright: All rights reserved.

D. Search Fingerprint

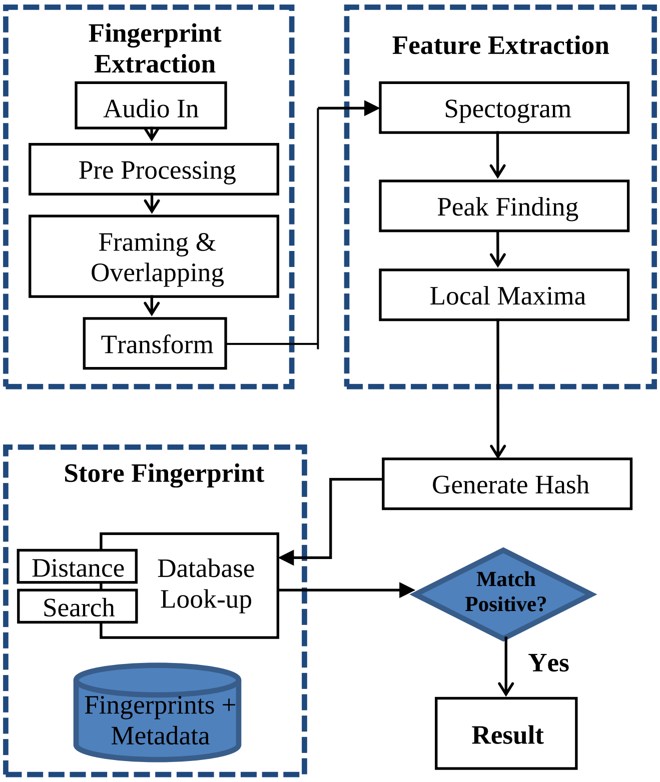

Fig. 1 Architecture of the Audio Copyright Infringement Detection System

Then the INDEX is included so that all of the queries will need to match giving a really quick retrieval to the proposed system. Next, the UNIQUE index just insures that there are no duplicates so that there is no need to waste space or no need to audio weight match by having underlying duplicates

The fingerprints are stored along with the regular track information. The regular track information includes the song ID, song name, file's hash value, offset value and match time. In the process of searching, it is assumed that the fingerprint object was already created.

When the user searches for the finger print, lookup functions is performed using the MySQL database in a distributed manner. It will return the matches by computing the track ranking and return the match list, thereby displays the respective matches of the audio fingerprint given as input.

The system has two use cases: one is to just fingerprinting the input song and storing them into a database in an effective manner, while the other is to detect the song information from the existing database, with a very small fragment of audio input of minimum 3 seconds.

V. Experimental Results

The proposed system is implemented using various python libraries like PyDub, NumPy and SciPy. The system's efficiency is experimented with a list of 45 songs and tested considering few essential parameters like speed and storage.

First, accuracy of the system is tested by feeding random segments of the same song of varying lengths. Some amount of noise was also added in by including a little indistinctive chatter while recording the audio input. The system was able to achieve 60 percent accuracy by randomly choosing from anywhere in the song for a single second. One extra second took the accuracy up to 96 percent.

To achieve 100 percent accuracy, the system needed a five-second segment randomly chosen anywhere from the song. To test its performance by speed, the system is tested with recording times and same are plotted using matched time. The accuracy analysis shown blow table.

Table I Results of Accuracy Analysis performed on a sample song

Number of Seconds	Number of Songs Predicted/Total Number of Songs	Accuracy (%)
1	35/60	58.33%
2	57/60	95%
3	58/60	96.65%
4	59/60	98.33%
5	60/60	100.0%
6	60/60	100.0%

Since the speed is mostly invariant of the particular song and more dependent on the length of the spectrogram created, it is tested on a single song, "Get Lucky" by Daft Punk.

From the illustrated graph, it is observed that the relationship is quite linear. The line is a least-squares linear regression fit to the data, with the corresponding line equation. The performance measure are shown in Table 2.The matching time includes recording time which leads to the fact that 3 times speed purely matches if the miniscule constant term is disregarded.

Table II Performance Measure

Time Recorded (s)	Time Recorded + Time to Match (%)
1	1.2
2	2.25
3	3.4
4	4.5
5	6
6	8
7	10
8	11
10	13
15	20
25	33
30	40

The round trip time is important for making matches. Hereforth, the system does not have to deal with the latency penalty in transferring fingerprint matches over the air at the local database instance. It provides the way to add round trip time to the constant term but does not affect the matching process.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)
© Copyright: All rights reserved.

VI. CONCLUSION AND FUTURE ENHANCEMENT

The proposed system will successfully detect copyright infringement of video and audio files by extracting feature points from the spectogram. The fingerprints are hashed into audio objects and stored into a database. While performing a look up onto the database, it searches in multiprocessing manner and returns an audio object which indicates the plagiarism of audio content. In future systems we propose to improvise the current algorithms in such a way that it supports video processing data and feature extraction algorithms can be improvised to identify the redundant audio segments and eliminate them from the hashed fingerprints.

REFERENCES

1. S. Chu, S. Narayanan and C. C. J. Kuo, "Environmental Sound Recognition With Time-Frequency Audio Features," in IEEE Transactions on Audio, Speech, and Language Processing, vol. 17, no.6, pp. 1142-1158, Aug. 2009.
2. P. Cano et al., "A review of Audio Fingerprinting", J.VLSI Signal Process., vol 41.,no.3 2005
3. R. Baeza-Yates and B. Ribeiro-Neto, Modern Information Retrieval, Addison Wesley, 1999.
4. S.Subramanya, R.Simba, B. Narahari, and A.Youssef, "Transform-based indexing of audio data for multimedia databases.", in Proc. Of Int. Conf. on Computational Intelligence and Multimedia Applications, New Delhi, India, Sept. 1999.
5. A.Kimura, K. Kashino, T. Kurozumi and H. Murase, "Very quick audio searching: introducing global pruning to the time-series active search" in Proc. Of Int. Conf. on Computational Intelligence and Multimedia Applications, Salt Lake City, Utah, May 2001.
6. Yu Liu, Hwan Sik Yun, and Nam Soo Kim, "Audio Fingerprinting Based on Multiple Hashing in DCT Domain", IEEE signal Processing letters, vol., 16,no,6,June 2009.
7. Priscilla, R., and M. Karthi. "Usage of Bioinformatic Data for Remote Authentication in Wireless Networks." ICTACT Journal on Image & Video Processing 9, no. 1 (2018).
8. Gladence LM, Ravi K, Karthi M. An Enhanced Method For Detecting Congestive Heart Failure -Automatic Classifier. Ramanathapuram: IEEE International Conference on Advanced Communication Control and Computing Technologies, ICACCCT-2014. 2014; p. 586-90
9. Shumeet Baluja and Michele Covell, " Content Fingerprinting Using Wavelets" in Proceedings of 3rd European conference on Visual Media Production(CVM), London UK, 2006,pp,198-207.
10. J.Haitsma, T. Kalker, and J. Oostven, "Robust audio hashing for content identification," in Procs. Of the International Workshop on Content-Based Multimedia Indexing, oct-2001.
11. willdrevo.com

AUTHORS PROFILE

Dr.R.Priscilla received her B.E degree in Computer Science and Engineering and received her M.E degree in Computer Science and Engineering from anna university, Chennai, india in 2005. She completed her PhD in anna university in the year 2013. She have nearly 18 years' experience in teaching profession. Currently she is working as a professor in St. Joseph's Institute of Technology,Chennai. Her research area include DataMining, Database Management System, Web Service and System Software.

Mr.Karthi.M, is working as Assistant Professor in the St. Joseph's Institute of Technology, Chennai. His teaching career spans over a period of 5 year and currently pursuing his Ph.D., in Anna University, Chennai. His area of interest is Data mining, Deep Learning, Design and Analysis of Algorithm. He has published 15 research papers in reputed journals and conferences. He has published a book titled as "Compiler Design" in the year of 2019. He was honored with "Gold medal" for his academic excellence in Master of Technology in Information Technology at Sathyabama University, Chennai 2015.

Mrs.C.Niroshini Infantia, M.E.,(Ph.D)., Assistant Professor has 6 years of work experience in the field of teaching and is currently working in St.Joseph's Institute of Technology, Chennai-600119. she was awarded my Bachelor of Engineering and Master of Engineering Degree in Computer Science Engineering from Anna University,Chennai and currently pursuing her Ph.D(CSE)., in Annamalai University. She have published few research papers in various scopus indexed journals and International Conferences. She is an active member in Computer Society of India and Institute of Engineers(IEI).

Neelambary V received M.Sc (Integrated Course) with a specialization of Computer Technology from St.Joseph's College of Engineering affiliated to Anna University in 2012 and M.E with a specialization of Software Engineering from St.Joseph's College of Engineering affiliated to Anna University in 2014. From 2014, she is working as an Assistant Professor at St. Joseph's Institute of Technology, Chennai. Her area of interest mainly focuses on software engineering and deep learning. She has attended and presented papers on various conferences and published journals on recent and emerging trends on Engineering and technology.

Lokeshwar Tabjula is a full stack developer at Verizon India. He did his bachelors in Information Technology at St.Joseph's Institute Of Technology and graduated in the year 2019. He has achieved the best outgoing student award in his batch. He is a FOSS enthusiast and has passion in teaching. He is an active volunteer at U&I India(an NGO to educate kids) and FSFTN(a community that practices opensource philosophy). He is also a toastmaster at Verizon's Toastmasters Club in Division O under district 82B.