ACRCloud Audio Fingerprinting System

Steve Wang
ACRCloud, Beijing, China
contact@acrcloud.com

ABSTRACT

ACRCloud [1] is a SELF-HELP cloud platform that helps companies and developers integrate Audio Fingerprinting technologies into their products. The underlying algorithm is audio fingerprinting. We take many commercial features into account during algorithm designing and developing. This short paper illustrates the designs behind the SW# submissions for MIREX 2015 Audio Fingerprinting task.

1. Introduction

As audio fingerprinting is getting wider employment in many application scenarios, such as music retrieval, audio and video recognition, radio stream monitoring, live TV detection, multi-screen interaction and so on, ACRCloud audio fingerprinting system is designed to be as compatible as possible. Commercial features we take into account include high recognition rate and fast speed, low memory occupation, online database updating, distributed deployment.

The data structure of ACRCloud audio fingerprint system is based on finding and matching local "landmarks" in a spectrogram as described by Wang [2]. The sample rate we use is 8kHz and the frequency bands we choose are segmented from 500Hz to 1500Hz with a logarithmic spacing.

The remaining sections discuss the considerations and solutions for commercial features mentioned before.

2. Recognition Rate and Speed

Because of the existence of super noisy query snippets in real world, deciding whether to do noise reduction is important during audio fingerprinting searching progress. We first analyse the average energy degree for the whole wav and separated frequency bands and then calculate a noisy degree variable (NDV) like mean square deviation. If the NDV exceeds a threshold predefined, a noise reduction is done before amplifying high frequency signals and reduce low frequency signals.

The threshold for selecting the "landmarks" also needs to be adjusted according to the NDV. Dynamic configuration for noise reduction and "landmarks" selecting helps saving computing resources a lot and keeps the recognition rate stable.

3. Memory Occupation

For ACRCloud audio fingerprinting system, Index compression is used for memory occupation.

As we all know, most audio fingerprinting use hash table for features indexing and searching. Fixed length key-value pair is normal. However, there'll be a lot of blank bits for such system. Maybe a key only occupy the first 8bits of an integer.

Index compression should helps reduce memory occupation. The most important is that

the compression algorithm needs to be rather simple and fast.

We select 4 optional key-value length and make the first 2bits for the flag of how long is the value after them.

This reduces almost 40% memory compared to non-compressing one and only increases 5% searching time.

4. Online Database Updating

The index structure is designed rather flexible because of online updating needs. The hash table composed with link-list and vectors which guarantees the scanning speed and adjustability at the same time.

5. Distributed Deployment

Large scale audio fingerprinting system should be distributed and we designed a distributed hash table for this system.

The values for each key is compact but different keys might be put in different nodes. And one key has several backups that put in different nodes with the original one just like Hadoop [3] do.

The deployment and adjustment of keys is done periodically and automatically to get better load balance. Distributed system helps balance computing resources, improve searching speed and avoid single node failure at the same time.

6. References

[1] ACRCloud, http://www.acrcloud.com
[2] A. Wang, “The Shazam music recognition service,” Comm. ACM, vol. 49, no. 8, pp. 44–48, Aug. 2006.
[3] Hadoop, http://hadoop.apache.org.