MICA
Music Identifying and Classifying Application

by

Jiaxi Li
Weichen Zhang
Yuan Fang

COMP5425
Multimedia Retrieval
Project Final Report

Project Duration: 6 Weeks
Project Start Date: 13/04/2016
Project End Date: 25/05/2016

Contents

1

Introduction

1.1. Motivation

The rapid growth in technology is greater than ever nowadays. In this context, music is getting more and more popular, and an increasing number of music has been infiltrated into lives of people and influencing them in different forms. With the development of the network and digitization technology, people are surrounded by music at all times. However, people may enjoy particular music without knowing what this song is or what it consists of. Therefore, identifying songs of interests becomes an executable and reasonable motivation for developing applications to suit this objective.

1.2. Project

Due the over-variety of music, this project is mainly focusing on Classical Music. There are three major functionality in this deliverable: match incoming music (through files and recordings) with our database, classify instrument played within music (through files and recordings), and rate the music played by yourself.

1.2.1. Searching

Similar to *Shazam*[1], we capture a piece of music clip from either microphone or unrecognized music file, and produce the best matching result in our database.

1.2.2. Instrument Classification

Beside the music name, what instrument played within music interests people. This project provides the ability of identifying instruments through either an unclassified file or microphone on you computer.

1.2.3. Rate Yourself

What more than matching and classifying is having the fun by practicing your own skills. Select the favorite song in the database, and upload personal plays (or through microphone). Then, score will be showing up, and parts that are not played well is pointed out to users on the screen.

---

[1]`http://www.shazam.com`

2

Project Specification

2.1. Group Work

The three major functionality, searching, instrument classifying and rating, are divided and assigned to Weichen Zhang, Jiaxi Li, Yuan Fang respectively. While everyone in the group is working individually with his part, GUI integration is conducted by all of the members regards to corresponding functionality.

2.2. Related Work

Burnett et al. mentioned in [2], "an audio fingerprint is a condensed digital summary, deterministically generated from an audio signal, and then can be used to identify audio samples or locate similar items in audio database in a very short time". With this determination of technique, Kim et al. then has proposed an algorithm for extracting music fingerprints directly from the audio signal in *IEEE International Conference* in 2008[7]. Moreover, she had concluded that this proposed algorithm outperformed the conventional state-of-art identification system. After that, there are various methodologies proposed to augment precision and reliability of audio fingerprints.

In 2013, Lee et al. [8] pointed out the inefficiency in Haitsma's fingerprint system[6]. The audio fingerprinting system provided by Haitsma uses a lookup table to identify the candidate songs in the database, which contains the sub-fingerprints of songs, and searches the candidates to find a song whose bit error rate is the lowest. However, the communication between system and database is extremely frequent which can result in the lost of sub-fingerprint duo to heavily degraded input signal. Lee, on the other had, solves this difficulties by reducing the number of database accesses using an partitioned method on each song.

In the last few years, audio fingerprint has been successfully applied and developed in repeating pattern detection[3], speech recognition [11], and audio content recognition[5].

For the purpose of recognizing instrument, Li et al. compared few approaches of extracting instrument within an audio, including convolutional neural networks (CNN) and Mel-frequency cepstral coefficients (MFCCs) [9].

MFCC computation technique is derived from a type of cepstral representation of the audio clip. It utilizes discrete cosine transform (DCT) for decorrelating log energies of filter bank output [10]. Then resulting log energies and filter banks can be used to calculating a set of coefficients that can best represent an audio. In fact, Bhalke et al. published a paper in *Digital Signal Processing Journal* in 2013 [1] and described the combination of higher order moments and MFCC improves recognition accuracy by more than 8% comparing to convention methodologies.

3

Problem Analysis

As Ganchev et al. mentioned in [4], different implementations has different performance.

4

Implementation

5

Discussion

5.1. Contribution
5.2. Future Work
5.3. Reflection on Preparing a Presentation

6
Conclusion

Bibliography

[1] DG Bhalke, CB Rama Rao, and DS Bormane. Musical instrument recognition using higher order moments. *Digital Signal Processing*, 5(4):133–138, 2013.

[2] Ian S Burnett, Fernando Pereira, Rik Van de Walle, and Rob Koenen. *The MPEG-21 book*. Wiley Online Library, 2006.

[3] Mei Chen, Qingmei Xiao, Kazuyuki Matsumoto, and Xin Luo. Detecting repeating pattern in fingerprint feature sequence. In *2015 Fifth International Conference on Instrumentation and Measurement, Computer, Communication and Control (IMCCC)*, pages 1864–1869. IEEE, 2015.

[4] Todor Ganchev, Nikos Fakotakis, and George Kokkinakis. Comparative evaluation of various mfcc implementations on the speaker verification task. In *Proceedings of the SPECOM*, volume 1, pages 191–194, 2005.

[5] Jeongseok Jo, Sungjun Han, and Jongweon Kim. Improvement of spectral fingerprint for audio content recognition. 2016.

[6] Antonius Adrianus Cornelis Maria Kalker and Jaap Andre Haitsma. Fingerprint database updating method, client and server, April 21 2009. US Patent 7,523,312.

[7] Samuel Kim, Erdem Unal, and Shrikanth Narayanan. Music fingerprint extraction for classical music cover song identification. In *Multimedia and Expo, 2008 IEEE International Conference on*, pages 1261–1264. IEEE, 2008.

[8] Sunhyung Lee, Dongsuk Yook, and Sukmoon Chang. An efficient audio fingerprint search algorithm for music retrieval. *Consumer Electronics, IEEE Transactions on*, 59(3):652–656, 2013.

[9] Peter Li, Jiyuan Qian, and Tian Wang. Automatic instrument recognition in polyphonic music using convolutional neural networks. *arXiv preprint arXiv:1511.05520*, 2015.

[10] Md Sahidullah and Goutam Saha. Design, analysis and experimental evaluation of block based transformation in mfcc computation for speaker recognition. *Speech Communication*, 54(4):543–565, 2012.

[11] Matthew Sharifi, Ben Shahshahani, and Dominik Roblek. Unified recognition of speech and music, December 29 2015. US Patent 9,224,385.