Music Plagiarism Detection Using Melody Databases

Jeong-Il Park[1], Sang-Wook Kim[2], and Miyoung Shin[3]

[1] Division of Computer, Information, and Communications Engineering
Kangwon National University, Chuncheon, Korea
inode777@empal.com

[2] College of Information and Communications
Hanyang University, Seoul, Korea
wook@hanyang.ac.kr

[3] School of Electrical Engineering and Computer Science
Kyungpook National University, Daegu, Korea
shinmy@knu.re.kr

Abstract. This paper addresses the development of a system that detects plagiarism based on similar melody searching. Similar melody searching is to find the melodies similar to a given query melody from a music database. For this purpose, we propose a novel similarity model that supports alignment as well as shifting. Also, we suggest a method for indexing the features extracted from every melody, and a method for processing plagiarism detection by using the index. With our plagiarism detection system, composers can easily search for the melodies similar to their ones from music databases. Through performance evaluation via a series of experiments, we show the effectiveness of our approach. The results reveal that our approach outperforms the sequential-scan-based one in speed up to around 31 times.

1 Introduction

Recently, the use of various types of multimedia data such as images, videos, and audios has shown its explosive growth, so the *content-based search* became of great importance[12][7][13][5][4][3]. For the successful content-based search, an indexing scheme and a query processing scheme are the key issues to be considered. Despite of its great advance, audio search has been less investigated than either image or video searches [8].

In this paper, we address the development of a system that detects plagiarism based on the *similar melody searching*, which is an operation that finds the melodies similar to a given query melody from a music database. Specifically, the plagiarism detection system is to examine if there exist such melodies in music databases as being similar to any melodies of a composer's interest. Without realizing it, any composers may be involved in diverse plagiarism disputes. The main purpose of this research is to help composers to avoid unnecessary plagiarism disputes by using the plagiarism detection system in advance.

Unlike such previous systems as [6][8][9][10], our plagiarism detection system has its unique characteristics as follows:

- A novel similarity model: it solves the problem of misjudgment by supporting alignment as well as shifting in the similarity model.
- Multidimensional indexing: it makes a basic framework for fast searching by employing a multidimensional index built on melody features.

R. Khosla et al. (Eds.): KES 2005, LNAI 3683, pp. 684–693, 2005.

- A three-step query processing: it provides fast search ability by taking a three-step query processing method, which consists of index searching, window stitching, and post-processing.

The results of performance evaluation show that our approach outperforms the sequential-scan-based one in the speed of searching up to around 31 times.

This paper is organized as follows. Section 2 describes a novel similarity model for computing the similarity between two different melodies. Section 3 discusses an indexing method for efficient similar melody searching from melody databases. Section 4 presents a three-step query processing method which employs the proposed indexing method. To show its superiority, we evaluate the performance of our system in Section 5. Finally, Section 6 summarizes and concludes the paper.

2 Similarity Model

This section describes a similarity model to compute the similarity between two different melodies.

2.1 Basic Model

A melody of a music is defined as a list of snatches, i.e. as a sequence $S = <(s_i, sL_i)>$ $(0<=i<n)$, called *a melody sequence*. Here, s_i denotes the i^{th} tone of a melody, and sL_i denotes the length of the i^{th} tone of a melody. Also, the number of tones, n, is called the length of a melody sequence. For example, the following melody with a C major key in four-quarters time is described as a melody sequence of length 4, i.e. <(sol, 1), (la, 1/2), (la, 1/2), (fa, 2)>[1].

For comparison of any two melody sequences $S = <(s_i, sL_i)>(0<=i<n)$ and $Q = <(q_j, qL_j)>(0<=j<m)$, we make the following assumptions:

- *Assumption 1:* S and Q should have the same meter. In other words, two melody sequences of different meters are excluded for similarity searching.
- *Assumption 2:* It always holds that n=m and $sL_i = qL_j$ if i=j, for all (i, j)'s. Thus, two melody sequences to be compared should be of the same length, and the lengths of their corresponding tones should be also the same.

For any two melody sequences S and Q satisfying these assumptions, the similarity between S and Q is computed based on Definition 1 described below.

Definition 1: For two melody sequences $S = <(s_i, sL_i)>(0<=i<n)$ and $Q = <(q_j, qL_j)>(0<=j<n)$, they are defined as being similar to each other if and only if the following condition, as well as Assumptions 1 and 2, are satisfied.

$$L_\infty(S, Q) = L_\infty(<s_0, s_1, ..., s_{n-1}>, <q_0, q_1, ..., q_{n-1}>) < \varepsilon$$

---

[1] For illustration, in this paper, a tone and its length are described as a symbol and a fraction, respectively. However, in practice, when being stored into standard MIDI file [11], both of them are described as integers, respectively.