Memory saving and real-time database search hardware acceleration based on temporal continuity of audio fingerprints

2110132 Kenshin Nomoto

In resent years, with the development of an environment in which anyone can easily access the Internet, the market for digital content via networks is expanding, replacing CDs and other media in the distribution of music and other content. Furthermore, the increasing distribution of music via networks has transformed the characteristics of that distribution. In addition to the one-way distribution from the record producer who owns the master recording rights to the consumer, which has been done through e-commerce sites, two-way distribution is now widely used, such as the exchange of content between users who were previously on the consumer's side. This environment of quicker access to information than ever before has been created by the spread of social media and distribution services, and is widely used in the distribution of content where the freshness of information, such as current affairs and trends, is important. While the distribution of of content using P2P file sharing software such as Winny and large-scale distribution platforms such as Youtube and Spotify has increased the accessibility and speed of content dissemination, there is a high demand for speed in identifying rights holders, despite the increased complexity of content distribution management in policing copyright on the digital information with the potential for reproduction.

Regarding these problems, methods have been proposed for copyright management using Fingerprint technology. Fingerprinting is a technology that identifies songs using a compact representation of thousands of bits by analyzing signal data, such as energy transitions in each frequency band of a song, and extracting features. A system for copyright management using fingerprints could be implemented in routers of Internet providers or origin servers of distribution platforms to active content use while reducing the burden on the user side. In this case, a fast feature extraction method using hardware in signal processing is known for the generation of fingerprints. However, there are many difficulties in realizing highly accurate and high-speed search for music search using fingerprints, due to the nature of fingerprints, in which a fingerprint generated from a quality-degraded sound source is not completely identical to a fingerprint generated from the original sound source, but is similar to it.

Since songs exchanged over the network are often compressed using lossy methods such as mp3 for the purpose of saving network bandwidth, transfer time, and storage space, when such a song is input to the system, it is

necessary to search for the nearest fingerprint in the database. In general, nearest neighbor element search has the problem of exploding computational complexity depending on the size of the database, known as the "curse of dimensionality". Thus, while music search using fingerprints has robustness against quality degradation due to compression and other factors, the complexity of the search method makes it difficult to increase the speed. In addition, the scale of data covered by copyright management systems ranged from several hundred thousand songs to several million songs or more for each subscription music distribution service, indicating that high-speed identification is required for large data that will continue to grow in the future.

To address these issues, a probabilistic but fast nearest neighbor fingerprint search method using LSH(Locality-Sensitive Hashing), called Staged LSH, has been proposed. In addition, it has been found that offloading the method to an FPGA can achieve higher speeds by taking advantage of multiple arithmetic units in hardware. On the other hand, existing methods sometimes required multiple hash tables in the system in order to raise the search accuracy to a certain level. In such cases, the hash tables occupies several times as much memory space as the fingerprint database, which greatly limits the size of the manageable music fingerprint database in environment with only small memory space, such as embedded devices.

First, we improve the method of search fingerprints generated from songs by focusing on the temporal continuity of fingerprints, thereby improving the performance in the trade-off between accuracy, search speed, and memory space efficiency. When implemented on an FPGA device with the same accuracy and with parameters selected with priority on memory saving, the evaluation results showed that the hash table was reduced by approximately 74.99% and the speed was increased by approximately 7.03 times compared to the existing method in a database of 4 million fingerprints. In addition, in Neighbor Staged LSH with improvements over the Staged LSH, we achieved a speedup of about 109.79 times when parameters with comparable accuracy and memory space efficiency were selected. The results show that the proposed method can be widely adapted to LSH-based fingerprint search methods. The above results show that our method, which focuses on the temporal continuity of fingerprints, maintains robustness against bit errors in fingerprint search, improves memory space efficiency, and is effective in speeding up fingerprint search.

Next, we proposed and discussed a parallelization strategy to speed up the hardware implementation do the method. For the existing parallelization method focusing on independence among hash tables, we analyzed the issues in using global memory as a storage location for major elements, and proposed a new content-oriented parallelization method with a data structure

based on content partitioning and a frame-oriented parallelization method that enables simultaneous search of multiple search elements. When each parallelization method was implemented and evaluated on a database of 4 million fingerprints for a small distortion fingerprint input, the 4-parallel content-oriented parallelization achieved a speedup of approximately 2.16 times faster than the 1-parallel one. In addition, we have confirmed that frame-oriented parallelization is effective in terms of search speed in environments where memory accessible in parallel is limited and the input fingerprints have low similarity.

As described above, while maintaining equivalent accuracy, we improved memory space efficiency and increased speed to raise the performance level, thereby extending the size of the music a certain amount of time and and the size of the music fingerprint database that can be stored in limited memory space.