

MP3 File Identification Based on Concurrence Order of Metadata

Youngseok Lee¹ and Jongweon Kim^{*}

¹*Department of Electronics, Chungwoon University
113, Sukgol-ro, Nam-gu, Incheon, Korea*

²*Department of Intellectual Property, Sangmyung University
20, Hongjimun-2gil, Jongno-gu, Seoul, Korea
yslee@chungwoon.ac.kr, jwkim@smu.ac.kr*

Abstract

In this paper we investigated applicability of metadata to identify music contents in MP3 file. Metadata in MP3 include information associated with music contents such as a table of Hoffman code, scale factors, stereo types.

We considered music contents in MP3 file as a function of metadata. Our idea is reasonable since music contents in MP3 file are controlled by characteristics of in metadata.

In order to identifying a MP3 file, we used a concurrence order of metadata. We assumed that MP3 files with other music contents have other concurrence orders in their metadata. In experiments, we extracted metadata from MP3 files and rearranged metadata as column vector of matrix. The rearranged metadata were analyzed various numerical analyses such as singular value decomposition.

1. Introduction

Music identification is the most typical application of audio searching technique. By comparing the characteristics of an unknown music query fragment. Usually transmitted from mobile phones on the wireless telecommunication network or from personal computers on the Internet, with those previously calculated and stored in characteristics database, matching results and related metadata are returned. Metadata in music file must represent music contents to differentiate from each other.

Currently, many researches have insisted on various methods for identifying music contents. Perhaps the most common method used is to identify the type or name of a file by the fingerprints in music contents [1]. This is an extremely unreliable method, as any user or application can change or delete fingerprints any time.

As a result, some users are able to conceal files from system administrators simply by renaming them to a filename with a different extension. While this doesn't conceal the existence of a file, it can conceal the nature of a file and can prevent it from being opened by the operating system.

Especially MP3 files take up most of music file in world music markets which are regardless of its legal or illegal. According to the recent report in [2, 3], about 90% of the music contents is traded by an MP3 file type in music market whether it is online or off line. Therefore the identification of MP3 file plays important role of a music searching, a storage indexing and a tracking of illegal distributions [4].

In this paper we investigated applicability of metadata to identify music contents in MP3 file. Metadata in MP3 include information associated with music contents such as a table of

*Corresponding Author

Huffman code, scale factors, stereo types. We considered music contents in MP3 file as a function of metadata. Our idea is reasonable since music contents in MP3 file are controlled by characteristics of in metadata.

The remained paper is organized as follows. Section 2 introduces the basic principles of MPEG Layer-3, bit stream data format and metadata structure. Section 3 details the steps of our experiment to identify MP3 file. Experimental results on occurrence order of metadata in a MP3 file are given in Section 4. Section 5 concludes this paper and point out possible ways of future research.

2. MP3 File Structure

A simplified illustration of MPEG Layer-3 encoder is

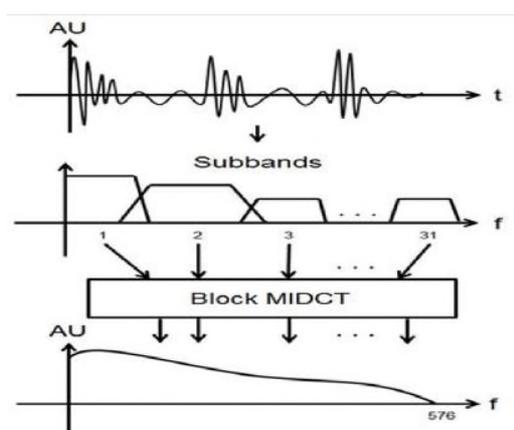


Figure 1. Simplified Frequency Transformation of MPEG Layer-3 Encoder

The input PCM signal is mapped into 32 equal bandwidth subbands through a polyphase filter bank, which simulate the critical bands in the human auditory system (HAS). The sub-band outputs are further subdivided by MDCT transform using long or short window to provide better spectral resolution. The long window allows greater frequency resolution for audio signals with stationary characteristics, while the short window provides better time resolution for transients.

Combined with other adjuvant techniques including psychoacoustic model, scale factor, Huffman coding, quantization etc., the final compressed bit stream is generated. Figure 1 displays the frame format of MP3 bit stream, and each frame has two granules to exploit further redundancies.

2.1. Overall Structure of MP3 File Format

Within an MPEG audio file, there is no main header, as an MPEG audio file is just built up from a succession of smaller parts called frames. Each frame is a data block with its own header and audio information as in Figure 2.

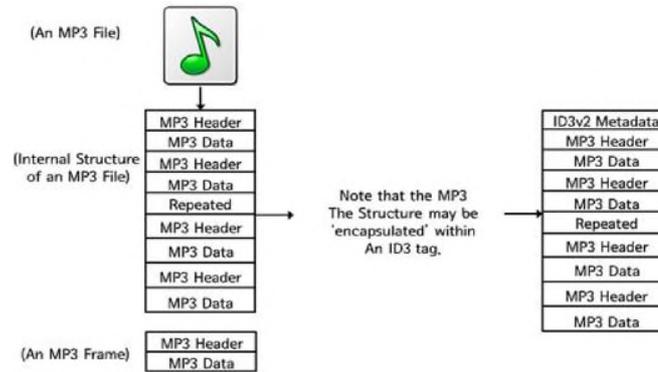


Figure 2. Simple Structure of MPEG Audio Layer-3 File

In the case of Layer I or Layer II, frames are totally independent from each other, so ONE can cut any part of an MPEG audio file and play it correctly. The player will then play the music starting from the first full valid frame it will find.

However, in the case of Layer III, frames are not always independent. Due to the possible use of the 'byte reservoir', which is a kind of internal buffer, frames are often dependent of each other. In the worst case, 9 input frames may be needed before being able to decode one single frame. If one needs to retrieve information about an MPEG audio file, you might simply locate the first frame, and retrieve information from its header. Information within other frames should be consistent with the first one, except for the bitrate, as you might be retrieving information from a variable bitrate (VBR) file.

In a VBR file, the bitrate can be changed in each frame. It can be used, as an example, to keep a constant sound quality during the whole file, by using more bits when the music is more complex and thus requires more bits to be encoded with a similar quality. The frame header including metadata itself is 32 bits (4 bytes) length.

The first twelve bits (or first eleven bits in the case of the MPEG 2.5 extension) of a frame header are always set to 1 and are called 'frame sync'. Frames may also feature an optional CRC checksum. It is 16 bits long and, if it exists, immediately follows the frame header. After the CRC comes the audio data. By re-calculating the CRC and comparing its value to the stored one, you can check if the frame has been altered during transmission of the bit stream.

2.2. Structure of Metadata in *sideinfo* of MP3 File

The side information part of the frame consists of information needed to encode the main data. The size depends on the encoded channel mode. If it is a single channel bit stream the size will be 17 bytes, if not, 32 bytes are allocated. The different parts of the side information are presented in Table 1 and described in detail below.

Table 1. Metadata in *sideinfo* of MPEG Audio Layer-3 File

Bit allocation	Single Channel	Dual Channel
Main_data_begin	9	9
Private_bits	5	3
Share	4	4+4
Information for granule		
Part2_3_length	12	12+12
Big_values	9	9+9
Global_gain	8	8+8

Scalefac compress	4	4+4
Window switching	1	1+1
For normal blocks		
Table_select	3*5	3*5+3*5
Region0_count	4	4+4
Region1 count	3	3+3
Preflag	1	2
Scalefac scale	1	2
Count1_table_select	1	2
Subtotal for first granule	59	118
Subtotal for second granule	59	118
Total number of bits	136	256
Total number of byte	17	32

The length of each field will be specified in parenthesis together with the fieldname above the actual description. If one length value is written the field size is constant. If two values are specified the first value will be used in mono mode and the second will be used for all other modes, thus these fields are of variable length. All tables below will assume a mono mode. The tables will change depending on mode since separate values are needed for each channel.

Using the layer III format there is a technique called the bit reservoir which enables the left over free space in the main data area of a frame to be used by consecutive frames. To be able to find where the main data of a certain frame begins the decoder has to read the main_data_begin value. The value is as a negative offset from the first byte of the synchronization word. Since it is 9 bits long it can point $(2^9-1) \times 8 = 4088$ bits. This means that data for one frame can be found several previous frames. Note that static parts of a frame like the header, which is always 32 bytes, are not included in the offset.

3. Proposed Method

The overall structure of proposed method is shown following as Figure 3.

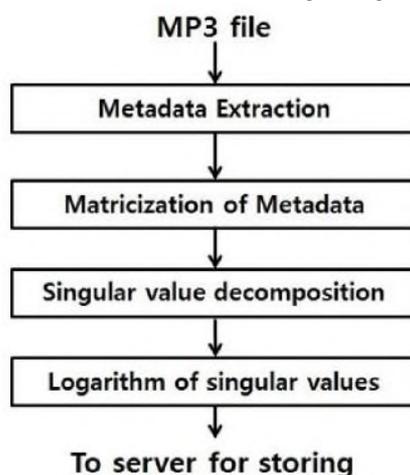


Figure 3. Overall Structure of Proposed Method

In Figure 3, we extracted the metadata from sideinfo blocks in MP3 file and then rearranged the metadata on matrix from at first. The rearranged metadata were analyzed singular value decomposition (SVD). And then logarithm of extracted singular values are calculated and stored in the underlying server to apply for identification of MP3 files.

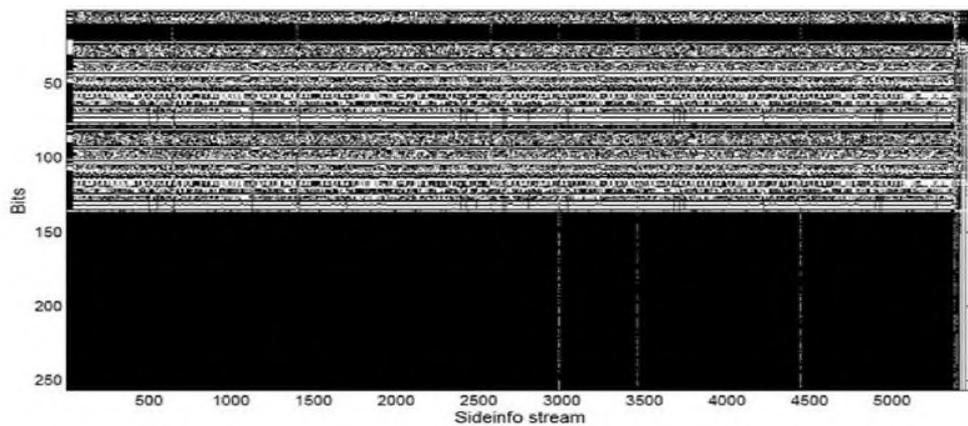
In our proposed method SVD plays an important role. SVD analyzes singular values of metadata matrix. SVD is a factorization of a real or complex matrix, with many useful applications in signal processing and statistics (VASIL KOLEV 2010). Formally, the singular value decomposition of an $m \times n$ real or complex matrix M is a factorization of the form.

$$M=U\Sigma V^* \quad (1)$$

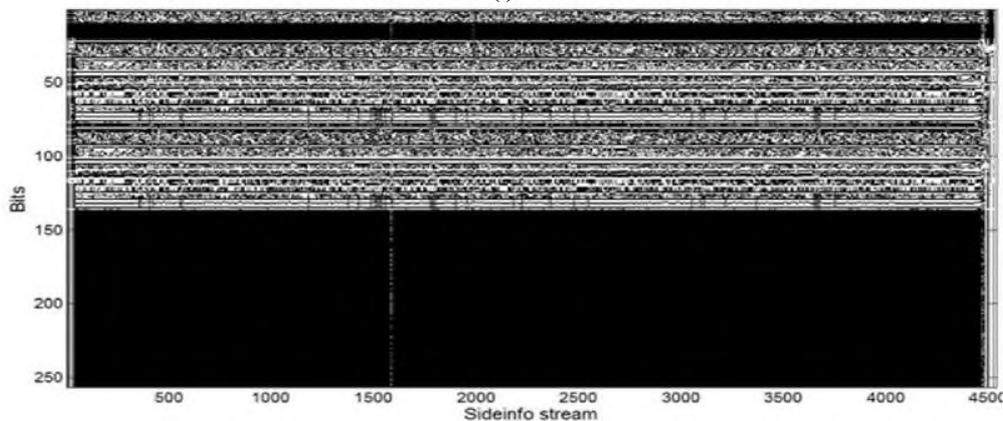
Where U is a $m \times m$ real or complex unitary matrix, Σ is an $m \times n$ rectangular diagonal matrix with nonnegative real numbers on the diagonal, and V^* (the conjugate transpose of V , or simply the transpose of ' V ' if ' V ' is real) is an $n \times n$ real or complex unitary matrix. The diagonal entries of Σ are known as the singular values of M . The m columns of U and the n columns of V are called the left-singular vectors and right-singular vectors of M , respectively.

4. Experimental Results

Author we conducted the experiment in accordance to the procedure of Figure 3. Firstly, the metadata were rearranged to a matrix form. Figure 5 shows the example of a rearranged metadata on a matrix form. The elements of metadata matrix are binary number, which are shown the bit '1' to white and bit '0' to black. In Figure 4, examples of two different MP3 files show mono and stereo type.



(a)



(a)

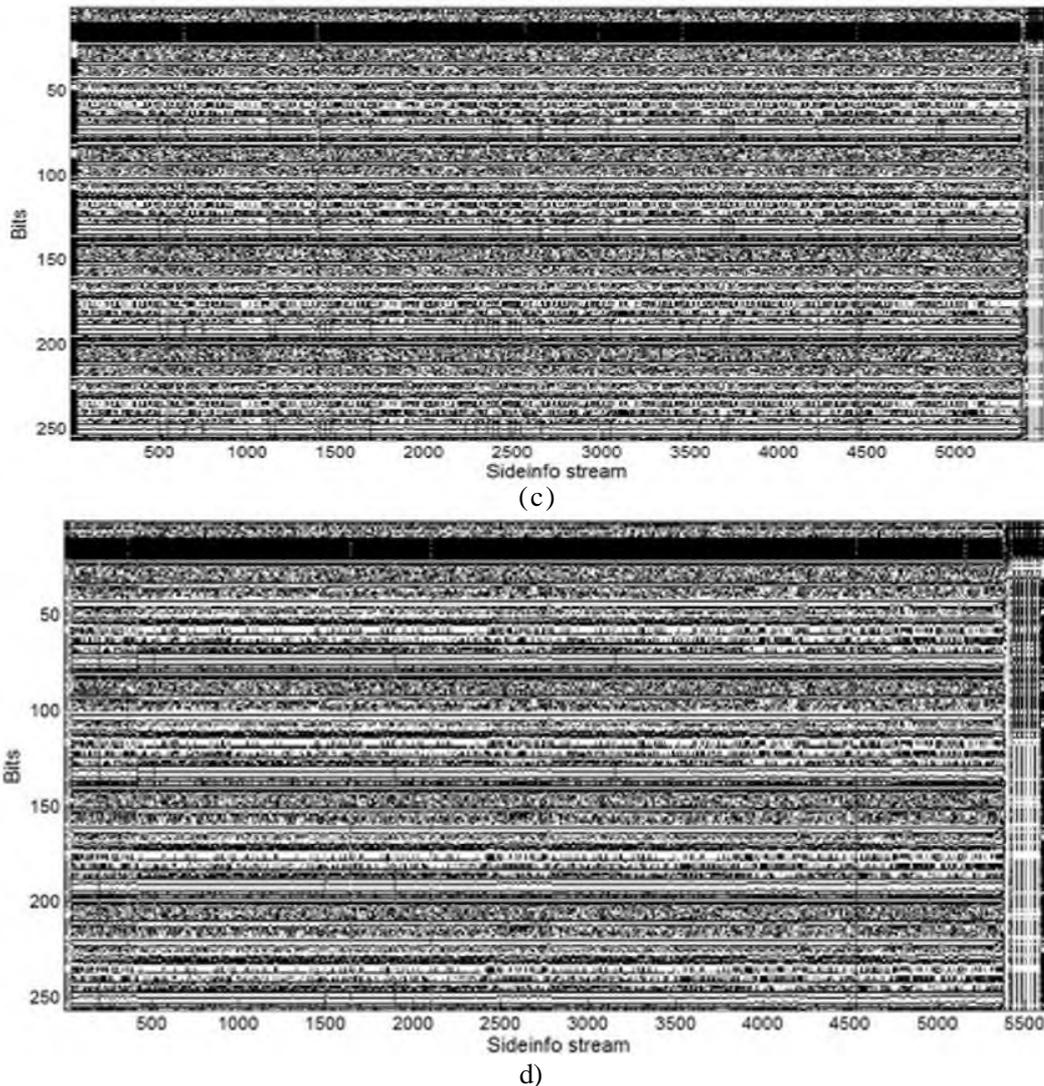
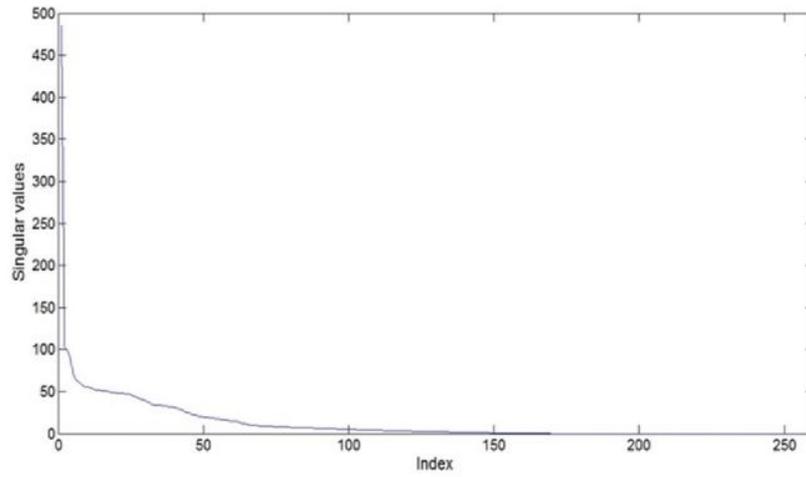


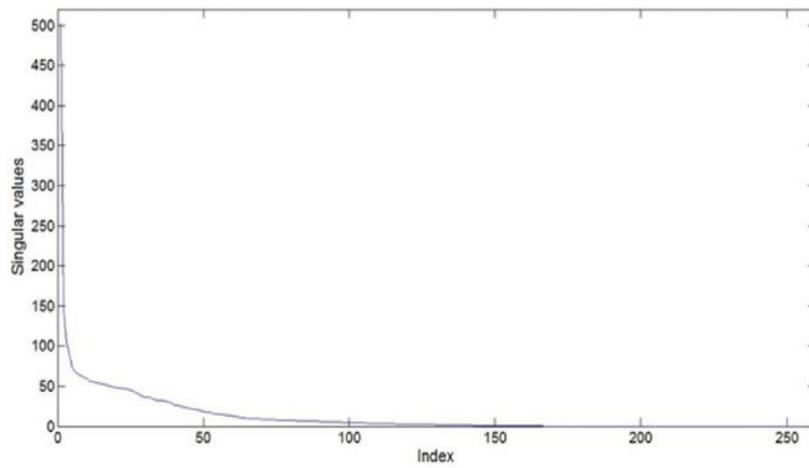
Figure 4. Two Example of a Rearranged Metadata on a Matrix Form: (a) and (b) are Mono Type, and (b) and (c) are Stereo Type

The number of column in metadata matrix is 256 and 136 in cases that the MP3 file is the stereo and mono type, respectively. The number of row depends on the number of metadata in the MP3 file. In MP3 files which used for extracting the metadata, the most MP3 file have metadata of a stereo type and some MP3 file are a mono type with the exception of some parts. Therefore the below of 156th bit in metadata matrix were bit '0' which were black color in Figure 4(a) and 4(b).

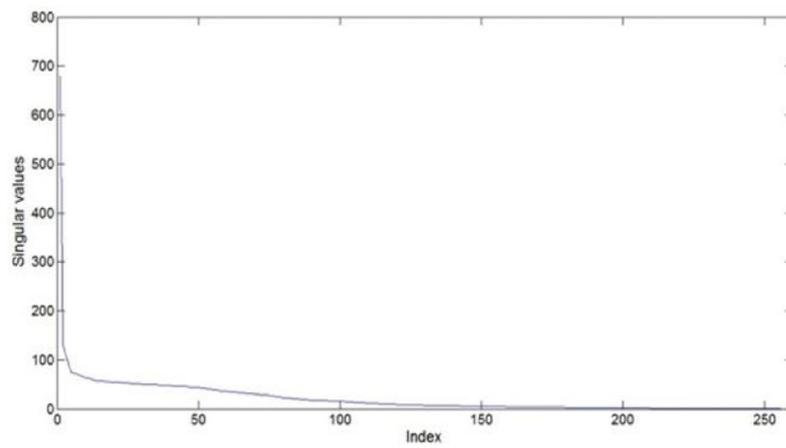
In SVD analysis of the metadata matrix, the metadata matrix was decomposed as in (1). All of 236 MP3 files were used to SVD analysis had full rank having 256 singular values. Figure 5 shows the curves of singular values by different MP3 files. In the figure, the horizontal axis is the index of metadata and the vertical axis is the linear scaled singular values.



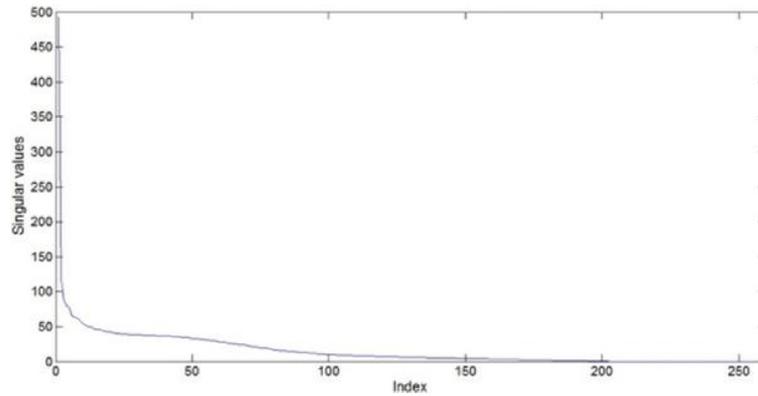
(a)



(b)



(c)



(d)

Figure 5. Curves of Singular Values for Metadata of Two MP3 File: (a) and (b) are Singular Values of Mono Type and (c) and (d) are Results of Singular Values of Stereo Type

Figure 6 shows the curves of singular values by different MP3 files. In the figure, the horizontal axis is the index of metadata and the vertical axis is the linear scaled singular values. The result of SVD analysis was related with some singular values Arranged in order of values. Value-ordered singular values that are dominant components in SVD, are different from each other MP3 file.

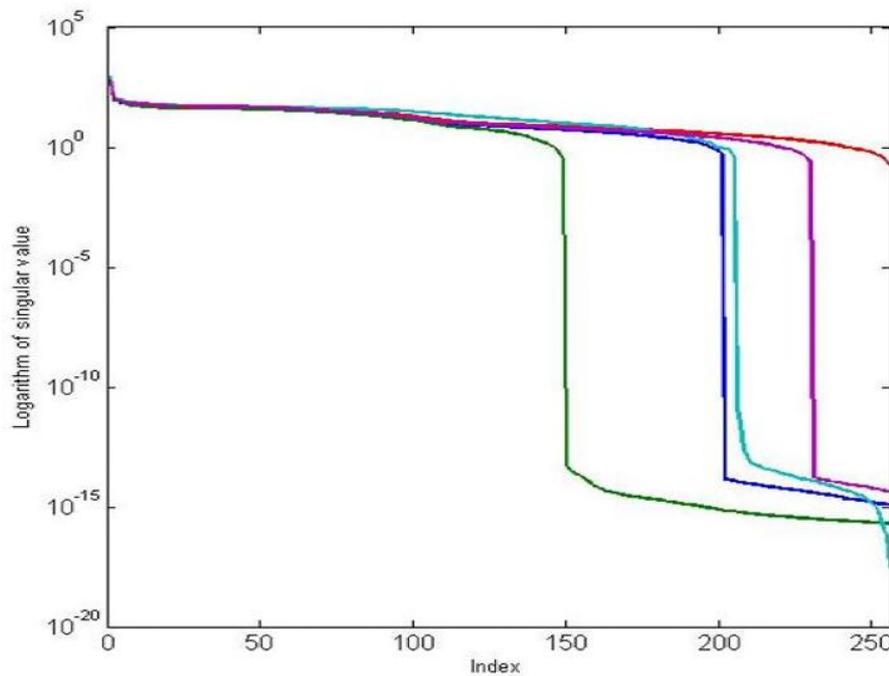


Figure 7. Curves of Singular Values of Rearranged Metadata by Different MP3 Files

Figure 7 shows the curves of singular values by different MP3 files from different scale perspective. In the figure, the horizontal axis is the index of metadata and the vertical axis is the logarithm scaled singular values. One can observe that the curves of logarithm- scaled singular values are clearly distinguishable. This result means that the logarithm- scaled singular values may be possibly used for the identification of MP3 file.

Especially, the interval from 100 to 10-15 of logarithm of singular values shows rapid changes in all MP3 files. In figure, blue and cyan colored curves represent mono- type MP3 files, and remaining curves are stereo type. When in extreme case that has the same as metadata among different MP3 files, our method cannot identify MP3 files. Since the metadata in underlying MP3 file are depend on the music contents in MP3 file format, Extreme case hardly occurs in the real case.

5. Conclusions

In this paper we investigated applicability of metadata to identify music contents in MP3 file. Metadata in MP3 include information associated with music contents. We observed that the logarithm scaled singular values can be used to identifying MP3 file format in binary data level. Especially, the interval from 100 to 10-15 of logarithm of singular values shows rapid changes in all MP3 files. Except for in extreme case that has the same as metadata among different MP3 files, our method shows the possibility to identify MP3 file format.

Acknowledgements

This research project was supported by Ministry of Culture, Sports and Tourism (MCST) and from Korea Copyright Commission in 2013.

References

- [1] P. Comon, X. Luciani and A. L. F. de Almeida, "Tensor Decompositions", Alternating Least Squares and other Tales. *Journal of Chemometrics*, vol. 23, (2009), p. 31.
- [2] M. M. McDaniel and M. H. Heydari, "Content Based File Type Detection Algorithms. In Proceedings of the 36th Hawaii International Conference on System Sciences", Hawaii, USA: IEEE Computer Society, (2002), p. 10.
- [3] V. Kolev, K. Tsvetkova and Milcho Tsvetkov, "Singular Value Decomposition of Images From Scanned Photographic Plates. Proceedings of the VII Bulgarian-Serbian Astronomical Conference", Chepelare, Bulgaria: Astron. Soc., (2010), p. 27.
- [4] W. Li, Yaduo Liu and X. Xue, "Robust Audio Identification for MP3 Popular Music", In SIGIR'10, Geneva, Switzerland: ACM, (2010), p. 8.

Authors



YoungSeok Lee, he received the Ph.D. degree from University of Seoul, major in signal processing in 1998. He is currently a professor of Dept. of Electronics at Chungwoon University in Korea since 1998. His research interests are in the area of image processing human visual system modeling and biomedical engineering.



JongWeon Kim, he received the Ph.D. degree from University of Seoul, major in signal processing in 1995. He is currently a professor of Dept. of Intellectual Property at Sangmyung University in Korea. He has a lot of practical experiences in the digital signal processing and copyright protection technology in the institutional, the industrial, and academic environments. His research interests are in the areas of copyright protection technology, digital rights management, digital watermarking, and digital forensic marking.