International Journal of Multimedia and Ubiquitous Engineering
Vol. 9, No. 12 (2014), pp. 307-320
http://dx.doi.org/10.14257/ijmue.2014.9.12.28

Audio Fragment Identification System

Xun Jin and Jongweon Kim

Dept. of Copyright Protection
Sangmyung University, Seoul, Korea
E-mail: jinxun@smu.ac.kr, jwkim@smu.ac.kr

Abstract

Audio identification technologies are becoming of increasing interest for copyright protection and digital forensic. In this paper, we propose an audio fragment identification system for efficient audio identifying applications in practice. To identify an audio file fragment, we have to decode it according to its format. Thus, we precede format classification of audio fragment using Support Vector Machine in advance. After identify the format, fragment recovery is implemented by adding a maximum similar frame header in front of the fragment. Then we extract a chroma feature from the decoded audio data to achieve audio identification. The experimental results show the evaluations of the format classification, fragment recovery and audio identification.

Keywords: *Audio Fragment Identification, Format Classification, Recovery, Support Vector Machine, Chroma Feature*

1. Introduction

With the rapid transmission of digital content and evolution of network technology, numerous problems about illegal distribution have been caused recently [1-5]. BitTorrent is a peer-to-peer file sharing technology which splits files into many fixed sized fragments for efficient distribution of files [6]. To prevent illegal distribution of digital audio through the BitTorrent, identification methods of audio fragments of MP3 and FLAC formats were proposed in [7, 8]. In [9, 10], audio identification method is also used in forensic technology under a hypothesis that the audio is correctly decoded.

In practical terms, most audio files are compressed or encoded while we are distributing or storing them. Therefore, first we have to decompress or decode the audio fragment, before identifying the audio. There are many kinds of compression methods, namely formats. In order to efficiently distinguish the format of fragment, we classify the formats of training audio fragment. The classification method used in our approach is Support Vector Machine (SVM). SVM is a powerful machine learning method because it is not limited by number of samples and dimensionality [11-14]. In [15-18], researchers used statistical features, such as mean, standard deviation, byte frequency distribution, Shannon entropy, N-gram and Hamming weight to classify the formats. Because of strong compression and entropy coding of audio file, it is hard to achieve high accuracy of classification only with statistical features. Therefore, we use patterns of sync words as a feature, which is the combination of statistical and structural features.

An audio fragment is obtained from randomly split audio file, which means it lost some audio or metadata. Because of many audio fragments failed to be decoded and output incorrect audio data, we apply a fragment recovery method to recover the audio. In order to increase the probability of successful recovery as much as possible, we add a maximum

ISSN: 1975-0080 IJMUE

similar frame header (MSFH) in front of the fragment. First of all, we collect many frame headers from audio samples. Then we rearrange the headers by the probability of occurrence of each property parameter in headers to form a header group. We search the group for MSFH which is concordant with the property of the fragment.

Several kinds of audio features can be extracted from audio data, such as chroma, rhythm, tonality and timbre [19]. The audio identification algorithm presented in this paper includes chroma extraction and two-dimensional (2D) cross-correlation. The entire spectrum of chroma feature is projected onto 12 bins which is representing the 12 distinct semitones of the musical octaves. Because the chroma feature is a powerful representation for audio data, it is suitable to be used to identify the audio fragment. The method of chroma extraction is based on that of [20], which involves the beat tracking. A beat tracker generates a beat-synchronous representation, which means there is one normalized feature vector per beat. The matching algorithm of chroma feature is 2D cross-correlation using 2D Fast Fourier Transformation (FFT). The identification performance of 12 chroma features using 2D correlation is much better than that of [20].

The overall system view of the proposed methods is shown in Figure 1. After we receive a fragment, the procedure of format feature extraction will be activated. Then the extracted feature will be classified by SVM classifier with the trained feature set. The procedure of fragment recovery search for the MSFH from the header group. After the fragment is decoded, chroma feature will be extracted by the procedure of audio feature extraction. Finally, the original audio will be detected by using 2D correlation.

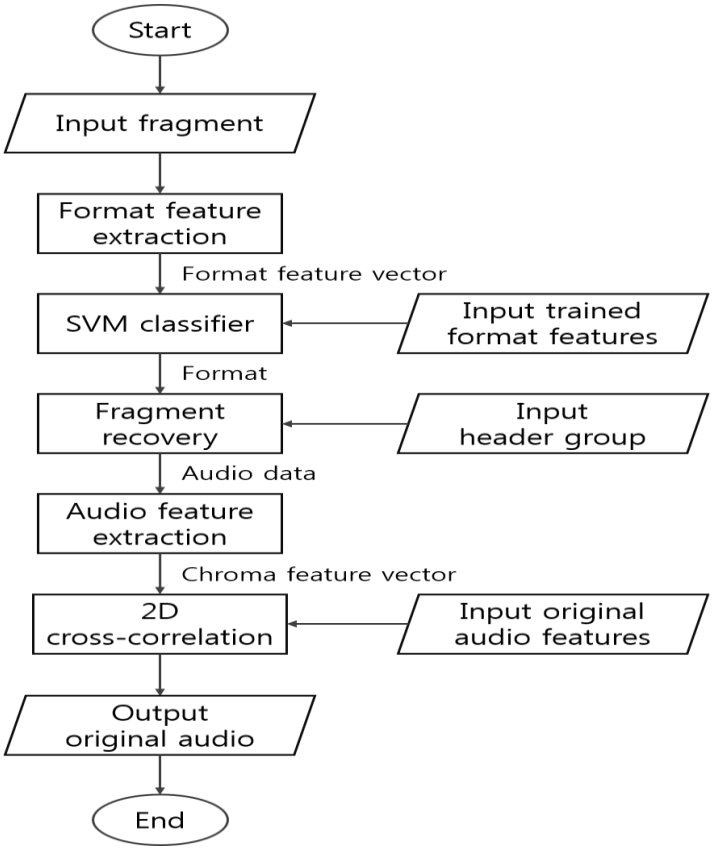


Figure 1. Overview of the Proposed System

The rest of this paper is organized as follows. In Section II, the method of format feature extraction and audio fragment format classification are presented. In Section III, we present the method of audio fragment recovery. In Section IV, the method of audio feature extraction and audio identification algorithm are presented. In Section V, we discuss the experimental results of format classification, fragment recovery and audio identification. In Section VI, we conclude the superiority of the proposed system.

2. Audio Fragment Format Classification

2.1. Format Feature Extraction

There are two types of pattern recognition approaches, namely structural and statistical approaches. The statistical patterns denote quantitative features such as mean, variance and frequency, whereas the structural patterns denote morphological features such as syntactic grammar and interrelationship [21]. Recently, many researches of format classification involve statistical approaches as in [15-18]. Because of the rapid growth of the capacity of multimedia, many formats utilize compression methods to reduce the cost and lead to generate high entropy data. Consequently, it is hard to achieve high accuracy of classification with statistical patterns. This approach proposes a hybrid pattern, which means the combination of the statistical and structural features.

There are several encoded audio frames in an audio fragment. Some other formats call them as packets, blocks or chunks. The frame of each format has its own sync word in front of the frame. The sync word of each format has its pattern. To play audio from any point of stream, the sync word is used to quickly and efficiently locate any positions of audio stream. For instance, the sync word of MP3 is '1111 1111 1111' and that of FLAC is '1111 1111 1111 10' in binary representation. In [22, 23], authors used the existence and quantity of sync words to classify the formats. But the sync words mentioned above can be appeared in any fragments of non-audio formats. To solve such a problem, we use the size information which indicates the length between two sync words.

If there are K audio formats needed to be classified, the set of entire sync words is defined as $S = \{s_1, s_2, \dots, s_K\}$, where the $s_k = \{bt_1, bt_2, \dots, bt_{M_k}\}, (k = 1,2, \dots K)$ denotes the sync word of kth format. The bt indicates one bit value and the M_k indicates the length of a s_k in bitwise. For instance, $M_k = 12$ in MP3. An audio fragment with a length of N is defined as $F = \{bt_1, bt_2, \dots, bt_N\}$. Decompose F into $N - M_k + 1$ subsets as follows:

$$F' = \{U_1, U_2, \dots, U_{N-M_k+1}\},\ U_c^k = \{bt_c, bt_{c+1}, \dots, bt_{c+M_k-1}\}, \quad (1)$$
$$c = 1,2, \dots, N - M_k + 1,$$

where U_c^k indicates cth subset while analyzing a s_k. The length and number of subsets are varied according to M_k. The positions $p_k = \{c_1^k, c_2^k, \dots, c_{I_k}^k\}$ where s_k occurred in F is defined as follows:

$$p_k = \{c | U_c^k = s_k, 1 \le k \le K\} \quad (2)$$

The I_k indicates the number of positions where match the s_k. We define the set of all the positions where S occurred in F as $P = \{p_1, p_2, \dots, p_K\}$. The positions of different k can be overlapped. The size information of each frame can be obtained from the parameters followed by s_k. For instance, the size information of MP3 is calculated with the padding bit, frequency and bit rate as in [24]. The function of calculating the length between two sync words at $c_i^k, (i = 1,2, \dots, I_k)$ and c_{i+1}^k is defined as $l_i^k = G_k(c_i^k)$, thus, a set of lengths

$L_k = \{l_1^k, l_2^k, \dots, l_{I_k}^k\}$ is obtained. If the word at c_i^k is equal to s_k after shifting l_{i-1}^k fromc_{i-1}^k, the word will be defined as available sync word. Hence, a set of positions of available sync words $C_k = \{c_1^k, c_2^k, \dots, c_{q_k}^k\}$ is obtained by using new positions $p_k' = L_k + p_k$ as follows:

$$C_k = \{c | p_k' \cap p_k, c \in p_k\} \quad (3)$$

The quantities of available sync words q_k are constructed to form a format feature vector $x = (q_1, q_2, \dots, q_K)$. Even though the number of audio formats is K, that of final classes is $K + 1$. The additional class indicates the non-audio formats. After extracting the x, a post-processing is applied to the x, which consists of linear interpolation and quantization, to obtain $x = (q_1, q_2, \dots, q_{K/r})$ where r is interpolation ratio.

2.2. SVM

The SVM proposed by Vapnik has attracted great interest in the research of machine learning. Because of the SVM is a very useful technique for data classification and provides better performance in terms of classification precision than other classification methods, it is used to classify the format features in this approach. It constructs two hyper planes called Plus-Plane (PP) and Minus-Plane (MP) along the support vectors as shown in Figure 2. The SVM finds an Optimal Separating Hyper plane (OSH) which can separate the feature vectors. This OSH has the largest margin between the other two hyper planes. If the margin is large, the result of classification is better. But most of the patterns are hard to classify precisely. To improve the performance of classification, the SVM maps feature vectors into a space of higher dimensions using kernel functions [13, 14].

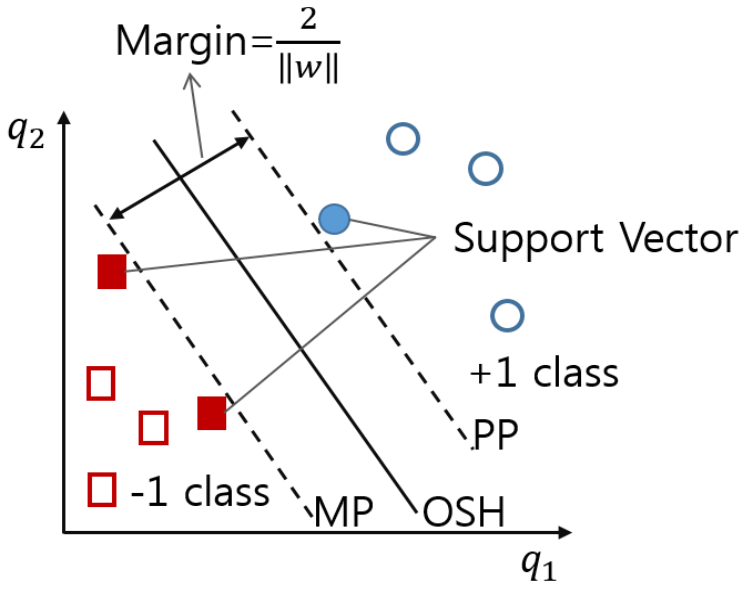


Figure 2. SVM

Defining w as the normal vector of the hyperplane, b as the bias and x as a feature point on the hyperplane, then the hyperplane is defined as follows:

$$\langle w, x \rangle + \mathrm{b} = 0 \quad (4)$$

Where the $\langle w, x\rangle$ indicates the inner product between w and x. The PP and MP are defined as follows:

$$\text{PP} = \{x|\langle w, x\rangle + b = +1\}$$
$$\text{MP} = \{x|\langle w, x\rangle + b = -1\} \quad (5)$$

The problem of maximizing the margin is called training of SVM. The margin is equal to $\frac{2}{\|w\|}$, thus, the problem is turned to the one minimizes reciprocal of the margin as follows:

$$\min\left[\frac{\|w\|^2}{2} + \mathcal{C}\sum_{i=1}^{K/r}\xi_i\right] \quad (6)$$

subject to

$$\langle w, x_i\rangle + b \geq 1 - \xi_i \;\; for \;\; y_i = 1$$
$$\langle w, x_i\rangle + b \leq -1 + \xi_i \;\; for \;\; y_i = -1$$
$$y_i(\langle w, x_i\rangle + b) \geq 1 - \xi_i,\; \xi_i \geq 0,\; \forall i = (1,2,\dots,K/r) \quad (7)$$

where y_i is the class to which the feature vector belongs, ξ_i is slack variables introduced to deal with misclassifications and $\mathcal{C}$ is trade-off parameter. The problem of optimization can be solved by the Lagrange dual problem as follows:

$$\max\left[\sum_{i=1}^{K/r}\alpha_i - \frac{1}{2}\sum_{i=1}^{K/r}\sum_{j=1}^{K/r}\alpha_i\alpha_j y_i y_j\langle x_i, x_j\rangle\right] \quad (8)$$

subject to

$$\sum_{i=1}^{K/r}\alpha_i y_i = 0,\; 0 \leq \alpha_i \leq \mathcal{C},\; \forall i = (1,2,\dots,K/r) \quad (9)$$

where α_i is Lagrange multiplier. Defining the maximized Lagrange multiplier as $\hat{\alpha}_i$ and the number of support vectors as N_{SV}, then the maximized weight vector is defined as follows:

$$\widehat{w} = \sum_{i=1}^{N_{SV}}\hat{\alpha}_i y_i x \quad (10)$$

Then the discriminant function is defined as follows:

$$f(x) = \text{sgn}\left[\sum_{i=1}^{K/r}\hat{\alpha}_i y_i\langle x, x_i\rangle + b\right] \quad (11)$$

Feature vectors are mapped to a space of higher dimensions by using a non-linear mappingΦ. The capability of the classification can be improved in this way. The SVM finds a separated hyper plane in that space with the maximal margin. Furthermore, the kernel function is defined as follows:

$$\kappa(x_i, x_j) = \langle\Phi(x_i), \Phi(x_j)\rangle \quad (12)$$

Researchers have been proposed many kinds of kernel functions. Here we just introduce some of them which are commonly used.

Linear kernel: $\kappa(x_i, x_j) = \langle x_i, x_j\rangle$

Polynomial kernel: $\kappa(x_i, x_j) = (\langle x_i, x_j\rangle + 1)^{\mathcal{p}}$

Radial Basis Function (RBF) kernel: $\kappa(x_i, x_j) = exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$

where the $\mathcal{p}$ and σ denote kernel parameters. The RBF kernel is used in this approach. Thus the final discriminant function is defined as follows:

$$f(x) = \text{sgn}\left[\sum_{i=1}^{K/r}\hat{\alpha}_i y_i\kappa(x, x_i) + b\right] \quad (13)$$

The SVM is designed to deal with binary problems, which means the labels of classes are two values: ± 1. In this paper, we need to classify $K+1$ classes. The additional class indicates the non-audio formats. Therefore, the binary problem is turned to a multiclass problem. To solve this problem, we split it into several binary classifiers. Splitting the problem into $K+1$ binary sub problems, for instance, class one to the other classes, class two to the other classes and *etc.*,

3. Fragment Recovery

An audio fragment is obtained from an audio file, which means it seriously lost some audio or metadata. It is hard to decompress or decode the fragment especially when we lost the metadata. If the size of the fragment gradually reduced, we cannot guarantee that there will be a frame header which contains the whole metadata. In order to maximize the probability of recovery, we use the MSFH.

We collected many headers corresponding to each format from the internet. The number of headers of kth format is defined as H_k. There are many types of parameters in a header, such as sample rate, bit rate and the number of channels. We define the number of parameters in a header of a format as D_k. A set of values which may occur in a parameter is defined as follows:

$$V_d^k = \left\{v_1^{(k,d)}, v_1^{(k,d)}, \dots, v_{J^{(k,d)}}^{(k,d)}\right\},\ d = 1,2,\dots,D_k \tag{14}$$

The number of headers which have values of $v_j^{(k,d)}$ in the dth parameter is defined as $\mathcal{N}_j^{(k,d)}$, then the maximum probability of the value occurrence in V_d^k is defined as follows:

$$\mathcal{P}_j^{(k,d)} = \frac{\mathcal{N}_j^{(k,d)}}{H_k},\ \mathcal{P}_{max}^{(k,d)} = \operatorname{argmax}\left\{\mathcal{P}_j^{(k,d)}, j \in \left[1, J^{(k,d)}\right]\right\} \tag{15}$$

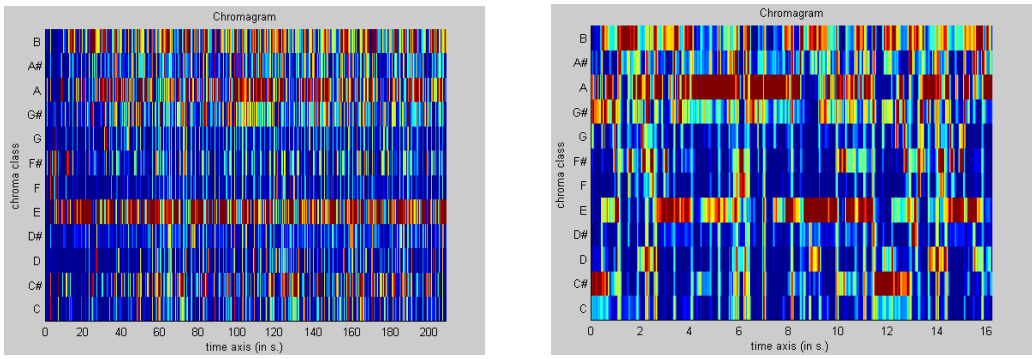
Then we select the $v_j^{(k,d)}$ which has the probability of $\mathcal{P}_{max}^{(k,d)}$ and rearrange the headers according to the descending order of $\mathcal{P}_{max}^{(k,d)}$, which means moving the headers with the values of the $v_j^{(k,d)}$ forward. After analyzing D_k parameters and D_k rearrangements, we remove the headers which have the same values of all parameters and use the rest of the headers to form a header group. According to the detected format of fragment, we search the corresponding group for the MSFH by decoding the added fragment. The header which can first make the fragment decodable is the MSFH.

4. Audio Identification

Since a beat is fundamental to the perception of music, beat tracking is an important procedure in computer simulation of audio signal. Although the audio components cannot be completely identified by people, we can track beats and keep time to music by foot-tapping. Thus, a computational model of beats can be built to track the beats [25, 26]. First, the audio signal converted into a time function at a low sampling rate. The strength of the onsets is reflected by taking the first order differences along the time. Summing across the frequency without the negative values. A high pass filter is used to remove the slowly changing DC offsets. To estimate a global tempo, auto-correlation is applied to the onset strength of the whole signal. The tempo is a pace reference which usually ranges from 40 to 260 beats per minute (BPM). After the beat tracker receives the best BPM, it tries to find optimized beat times.

We can extract useful audio information from the chroma distribution. The chroma features is composed of 12 dimensional vectors. Each dimension represents the intensity

associated with a semitone. Each feature vector is recorded per beat. The 12 elements of a feature vector capture the broad harmonic accompaniment and the dominant note. Using the phase derivative within each FFT bin to get a better resolution estimation of underlying frequency and identify strong tonal components within spectrum. Figure 3 shows chromagrams of an audio file and one fragment of it.



(a) Chromagram of an audio file (b) Chromagram of one fragment

Figure 3. Chromagrams of an Audio File and One Fragment of It

The conventional matching algorithm takes cross-correlation each row vector of the chroma feature matrix and calculates the summation of the correlation coefficients. Then it shifts the positions of row vectors to the next and takes cross-correlation again. After 12 rounds of shifts, it normalizes the correlation coefficients and selects the maximum peak from the coefficients as shown in Figure 4. The density distribution of some row vectors of an audio fragment may be close to that of other audio. It means the correlation between two chroma row vectors of different audio will be increased, when one of them shifts and leads to errors. Therefore, we propose an improved method to avoid such misidentifications. We combine the 12 one-dimensional (1D) chroma bins to generate a 2D chroma pattern. It reduces the similarity influence of 1D chroma row vectors on correlation and obtains higher accuracy of identification. We can raise the matching speed and reduce the complexity by using the 2D FFT to compute 2D correlation. Finally, the audio with the max value among the maximum peaks is identified to be the original audio.

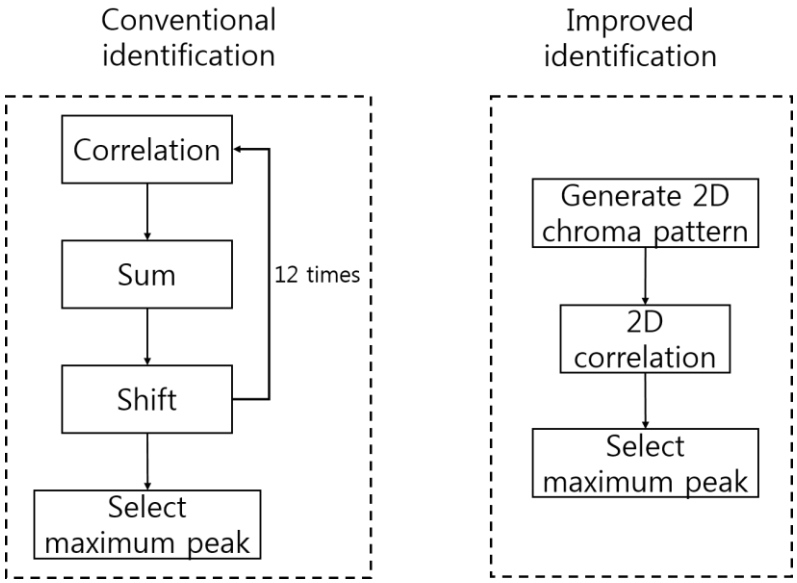


Figure 4. Structures of Conventional and Improved Methods

5. Performance Evaluation

The evaluation is composed of three types of experiments. The first one is to evaluate the classification error rate of the fragment format classification. The second one evaluates the probability of recovering the audio fragments. The last one evaluates the probability of identifying the audio fragments. All the fragments used in the experiments contain no file headers or footers. Most of the fragments were split from the song files including folk, country, rock, classical and hiphop.

In the experiment of fragment format classification, we evaluated three audio formats: MP3, AAC and FLAC. The training samples consist of 250 fragments randomly extracted from 250 files. The test samples consist of 100 fragments randomly extracted from other 100 files. The percentage of samples corresponding to each audio format in the sample space is 25%. The rest of samples are non-audio format fragments, such as EXE, ZIP, PDF, and *etc.*, The Table 1 shows the classification error rates of three different classification algorithms with different lengths of fragments. The classification performance of SVM is better than those of the other two algorithms. The probability of classification is greater than 90% when the length of fragment is larger than 128 KB.

Table 1. Classification Error Rates of Three Different Classification Algorithms

Fragment length	SVM	Decision Tree	Linear Discriminant Analysis
128 KB	13.89%	14.81%	15.74%
256 KB	8.33%	9.26%	12.04%
512 KB	4.63%	6.48%	10.19%
1 MB	3.7%	3.7%	6.48%

In the experiment of fragment recovery, we evaluated the fragments with the lengths from 128 KB to 16MB. The Figure 5 shows the probabilities of successfully decoding with 350 FLAC audio fragments. As it can be seen from the figure, the probability is growing tendency with the increasing length of fragment and exceeds 90%.

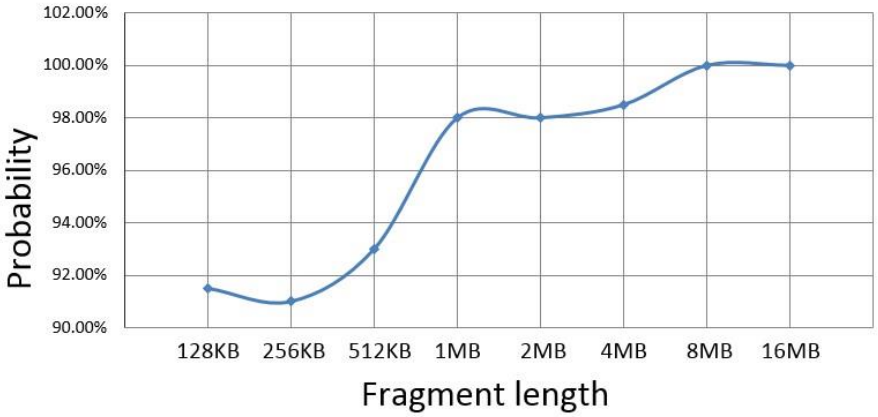
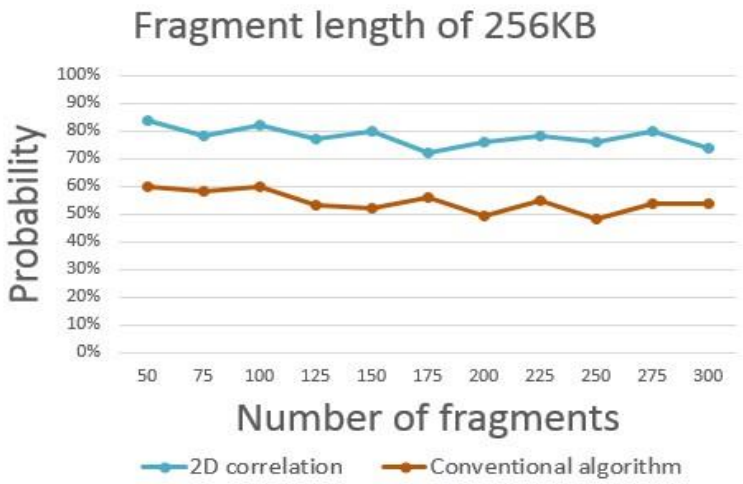
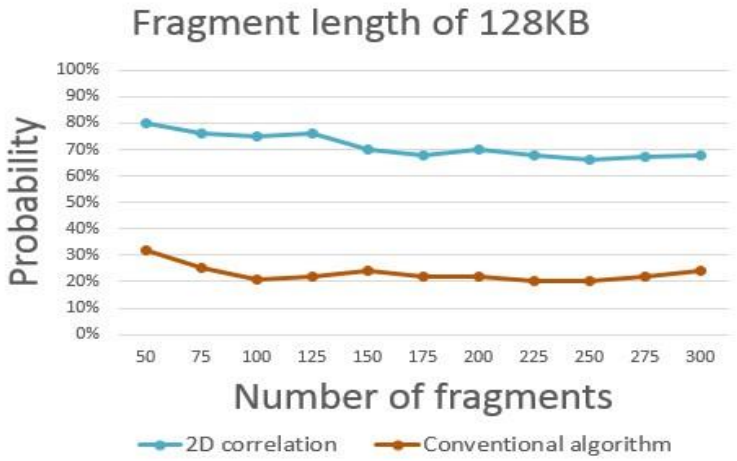


Figure 5. Probabilities of Successfully Decoding

In the experiment of audio identification, we evaluated the audio with the recovered fragments. The experiment was conducted on the goals including: (1) the accuracy of the identification for the proposed method with different lengths of fragments, (2) the comparisons between the proposed method and the conventional method, (3) the robustness of the proposed method against recompression attack. Figure 6 shows the probabilities of successful identification with the increasing number of fragments. As it can be seen in the Figure 6 (a), the average probability of the proposed method is greater than that of conventional method about 20% with 256 KB fragments, whereas in Figure 6 (b), it is greater than that of conventional method about 40% with 128 KB fragments.



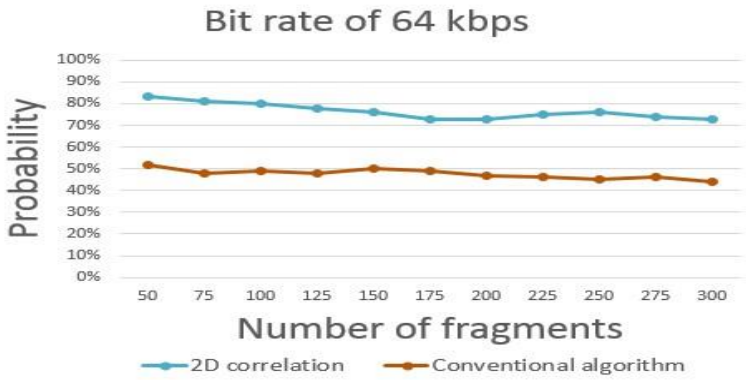
(a) Probabilities of successful identification with 256 KB fragments



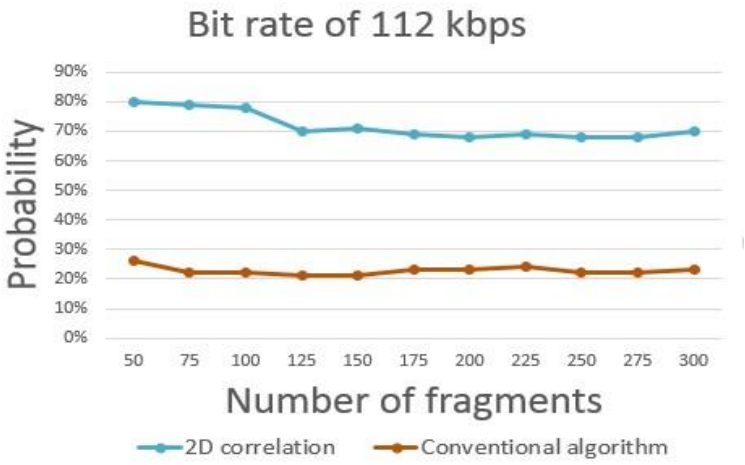
(b) Probabilities of successful identification with 128 KB fragments

Figure 6. Probabilities of Successful Identification

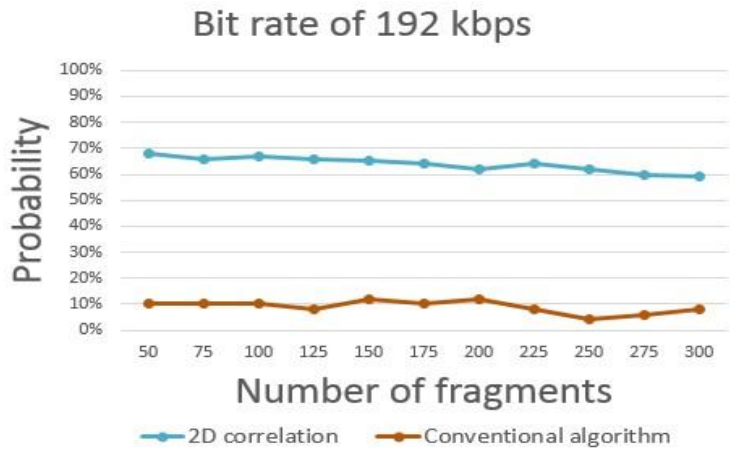
In practice, the audio files may be distributed by internet users after recompression. Therefore, we evaluated the robustness of the proposed method after recompression attack with bit rate of 64 kbps, 112 kbps and 192 kbps. Figure 7 shows the probabilities of successful identification with 128 KB fragments. As it can be seen in the Figures, the probabilities of conventional method are unstable, whereas those of the proposed method maintain at around 80%.



(a) Probabilities of Successful Identification with Bitrate of 64 kbps



(b) Probabilities of successful identification with bitrate of 112 kbps



(c) Probabilities of successful identification with bitrate of 192 kbps

Figure 7. Probabilities of Successful Identification with Different Compression Strengths

6. Conclusion

In this paper, an audio fragment identification system is proposed for practical application. Because most audio data are compressed or encoded when they are being distributed or stored, we have to decompress or decode them according to their format before identifying the audio. In order to efficiently distinguish the format of a fragment, we proposed format classification of audio fragments using SVM. After the format is identified, we implement fragment recovery procedure by adding a maximum similar frame header in front of the fragment. Then we extract chroma features from the recovered audio data and identify it by using 2D cross-correlation. The experimental results show the precision of classification is greater than 90% when the length of fragment is larger than 128 KB. The average probability of successful identification of improved method is greater than that of conventional method about 40% with 128 KB fragments.

Acknowledgments

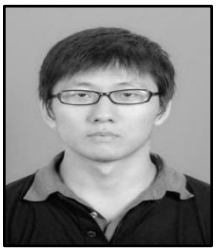
This research project was supported by the Ministry of Culture, Sports and Tourism(MCST) and the Korea Copyright Commission in 2014.

References

[1]. J. Kim, N. Kim, D. Lee, S. Park and S. Lee, "Watermarking two dimensional data object identifier for authenticated distribution of digital multimedia contents", Signal Processing: Image Communication, vol. 25, **(2010)**, pp. 559-576.
[2]. Y. Lee and J. Kim, "Robust Blind Watermarking scheme for Digital Images Based on Discrete Fractional Random Transform", Communications in Computer and Information Science, vol. 263, **(2011)**, pp. 139-145.
[3]. J. Nah, J. Kim and J. Kim, "Video Forensic Marking Algorithm using Peak Position Modulation", Applied Mathematics & Information Sciences, vol. 7, no. 6, **(2013)**, pp. 2391-2396.
[4]. J. Lee and J. Kim, "Modeling of a Copyright Protection System for the BitTorrent Environment", International Conference on Computer Applications for Security, Control and System Engineering, CCIS339, **(2012)**, pp. 46-53.
[5]. M. I. H. Sarker, M. I. Khan, K. Deb and M. F. Faruque, "FFT-Based Audio Watermarking Method with a Gray Image for Copyright Protection", International Journal of Advanced Science and Technology, vol. 47, **(2012)**, pp. 65-76.
[6]. P. Sharma, A. Bhakuni and R. Kaushal, "Performance Analysis of BitTorrent Protocol", Communications NCC, National Conference, **(2011)**, pp. 1-5.
[7]. X. Jin and J. Kim, "Partial Identification Analysis for MP3 Music", Journal of Advances in Computer Networks, vol. 2, no. 2, **(2014)**, pp. 151-154.
[8]. R. Jin and J. Kim, "Analysis of FLAC Music Pieces Recovery", Journal of Advances in Computer Networks vol. 2, no. 2, **(2014)**, pp. 134-137.
[9]. R. Maher, "Audio forensic examination", Signal Processing Magazine, IEEE, vol. 26, **(2009)**, pp. 84-94.
[10]. A. J. Cooper, "Detection of Copies of Digital Audio Recordings for Forensic Purposes", Open University, BLDSC, **(2006).**
[11]. V. Vapnik, "The Nature of Statistical Learning Theory", Springer-Verlag, **(1995).**
[12]. B. E. Boser, I. Guyon and V. Vapnik, "A training algorithm for optimal margin classifiers", In Proceedings of the Fifth Annual Workshop on Computational Learning Theory, **(1992)**, pp. 144-152.
[13]. D. K. Srivastava and L. Bhambhu, "Data classification using support vector machine, Journal of Theoretical and Applied Information Technology, vol. 12, **(2010).**
[14]. M. Fauvel, J. Chanussot and J. A. Benediktsson, "A Combined Support Vector Machines Classification Based on Decision Fusion", Geoscience and Remote Sensing Symposium, IEEE International Conference, Denver, CO, USA, **(2006).**
[15]. S. Fitzgerald, G. Mathews, C. Morris and O. Zhulyn, "Using NLP techniques for file fragment classification," Digital Investigation, vol. 9, **(2012)**, pp. 44-49.
[16]. W. C. Calhoun and D. Coles, "Predicting the types of file fragments," Digital Investigation, vol. 5, **(2008)**, pp. 14-20.
[17]. G. Conti, S. Bratus, A. Shubina, B. Sangster, R. Ragsdale, M. Supan, A. Lichtenberg and R. Perez-Alemany, "Automated mapping of large binary objects using primitive fragment type classification," Digital Investigation, vol. 7, **(2010)**, pp. 3-12.
[18]. M. C. Amirani, M. Toorani and A. A. Beheshti, "A New Approach to Content-based File Type Detection," Proceedings - IEEE Symposium on Computers and Communications, **(2008)**, pp. 1103-1108.
[19]. O. Lartillot and P. Toiviainen, "A Matlab Toolbox for Musical Feature Extraction from Audio," Proc.of the 10th Int. Conference on Digital Audio Effects, Bordeaux, France, **(2007)**, pp. 10-15.
[20]. D. P. W. Ellis and G. E. Poliner, "Identifying 'Cover Songs' with Chroma Features and Dynamic Programming Beat Tracking," Acoustics, Speech and Signal Processing, ICASSP, presented at the IEEE International Conference, **(2007)**, pp. 1429-1432.
[21]. R. T. Olszewski, "Generalized Feature Extraction for Structural Pattern Recognition in Time-Series Data Thesis Committee," Engineering Education, vol. 35, **(2001)**, pp. 1-125.
[22]. V. Roussev and C. Quates, "File fragment encoding classification-An empirical approach", Digital Investigation, vol. 10, **(2013).**
[23]. V. Roussev and S. L, Garfinkel, "File Fragment Classification-The Case for Specialized Approaches", Systematic Approaches to Digital Forensic Engineering, Fourth International IEEE Workshop, Berkeley, CA, **(2009)**, pp. 3-14.
[24]. E. Kalpana, V. Sridhar and M. R. Prasad, "MPEG-1/2 audio layer-3(MP3) ON THE RISC based ARM PROCESSOR (ARM92SAM9263)", International Journal of Computer Science Engineering, vol. 1, **(2012).**

[25]. M. Goto and Y. Muraoka, "A Real-time Beat Tracking System for Audio Signals," International Computer Music Association, ICMC, **(1995)**, pp. 171-174.
[26]. T. Jehan, "Creating Music by Listening," PhD thesis, MIT Media Lab, Cambridge, MA, **(2005).**

Authors



Xun Jin, he received his B.S. degree in Computer Science and Technology from Fujian Agriculture and Forestry University, China, in 2011. He is currently pursuing the Ph.D. degree in Copyright Protection, Sangmyung University, Korea. His research interests are digital image/video watermarking, multimedia forensics, pattern recognition, digital signal processing, and information security.

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 at the institute, the industry, and College. His research interests are in the areas of copyright protection technology, digital rights management, digital watermarking, and digital forensic marking.