COMPRESSION HISTORY IDENTIFICATION FOR DIGITAL AUDIO SIGNAL

Da Luo[‡], *Weiqi Luo*[†], *Rui Yang*[‡], *Jiwu Huang*[‡]

[†] School of Software, Sun Yat-Sen Univ., Guangzhou, P.R. China, 510006
[‡] School of Information Science and Technology, Sun Yat-Sen Univ., Guangzhou, P.R. China, 510006

ABSTRACT

Compression history identification plays a very important role in digital multimedia forensics. However, most existing literatures mainly focus on digital image forensics, and just a few works consider digital audio. In this paper, we investigate two popular compression schemes in digital audio, that is, MP3 and WMA, and try to reveal the compression history for a questionable audio signal in the original uncompressed WAV format via analyzing some statistical characteristics of the modified discrete cosine transform coefficients of the audio. The extensive experimental results have shown that the proposed method can effectively identify whether the given audio has been previously compressed with MP3 and/or WMA, and can further estimate the hidden compression rates, even the compression rate is as high as 128 K bps(bits per second).

Index Terms— Audio compression history, MDCT coefficients, Compression rate estimation

1. INTRODUCTION

With sophisticated editing technologies, it is becoming increasingly easy to tamper digital media without leaving any obvious visual or auditive clues even for our common users. Many problems related to serious moral, ethical and legal consequences have occurred in our real life. Nowadays, digital multimedia authentication faces the challenge.

Digital watermarking and digital signature are the two typical technologies for digital multimedia authentication. However, both technologies need some additional side information such as digital watermark or signature at the time of detection. In many real applications, however, these methods will become useless since it is impossible to extract any available side information from the questionable data.

Multimedia forensics is an emerging research field of information security in the last several years. This technology doesn't need any side information during detection. By analyzing the inherent features within the multimedia, it can provide forensics information on how multimedia data is acquired and processed. Up to now, many effective methods have been proposed for digital image forensics. However, just a few literatures have been reported for digital audio.

The existing literatures related to digital audio forensics mainly concentrated on the following aspects. The works [1, 2, 3] have addressed the problem of determining whether a digital signal has been cut/spliced or not. Some researches [4, 5] reported how to determine the used microphones and the environments of recorded audio. Our previous work [6] uses the numbers of small MDCT coefficients as features to discriminate fake-quality MP3 format files from normal ones. Qiao *et al.* [7] enhanced the above method [6] to effectively detect double compression including up-transcoding and down-transcoding in MP3 files. Feng and Doërr [8] proposed a method to detect the re-compression of speech signals between the different quantization strategies.

In many applications, compression history identification is one of important issues in digital multimedia forensics. In this paper, we try to determine whether a given audio signal in the original uncompressed WAV format has been previously compressed or not, and then further to estimate the compression type (*i.e.* MP3 and WMA) and the corresponding compression rate used during compression. To our best knowledge, no relative works have been reported previously. Please note that one potential method [3] for compression identification has been included in our comparative experiments, see Section 3. Another important reason why we are interested in such issue is that, piracy compact disks (CD) can be burned from the decompressed MP3 or WMA file format music. Such CD is low quality in essence because it is transformed from those low bit-rate audio. So it is necessary to develop a technology to uncover the compression history of an audio to reveal such fake-quality music.

In this paper, the two popular compression schemes in audio compression (MP3 and WMA) have been investigated. Inspired by our previous work [9] in JPEG history estimation for digital images, we found that the quantization artifacts detection is the crucial issue in compression history estimation. In the next Section 2, therefore, we will mainly investigate the modified discrete cosine transform (MDCT) coefficients of the audio signal, namely, those coefficients before the quantization operation during MP3 compression[1]. And then propose a 21-D feature vector to represent the statistical characters of the coefficients at different compression rates. The extensive experimental results shown in Section 3 have demonstrated the effectiveness of the proposed method for estimating the compression history of the MP3 and WMA compression schemes, and their combinations.

The rest of the paper is organized as follows. Section 2 briefly reviews the MP3 coding, and proposes the statistical features to expose the quantization artifacts introduced by the MP3 compression. Section 3 shows the experimental results and analysis both for MP3 and WMA audio. The concluding remarks and future works will be given in Section 4.

This work is supported by the 973 Program (2011CB302204), NSFC (U1135001,61003243), the funding of Zhujiang Science & technology (2011J2200091) and China Postdoctoral Science Foundation (20080440795). J. Huang is also with the State Key Laboratory of Information Security, Beijing 100049, China.

[1] Please note that there is no literatures or technical reports about the compression details of the WMA data. Therefore, the proposed feature vector in this paper is mainly derived from the analysis on MP3 compression.

2. PROPOSED METHOD

Similar to most lossy compression schemes in digital image and video, such as JPEG and H.264, quantization operation step in MP3 compression will introduce some quantization artifacts in the corresponding frequency coefficients and thus indicate the compression history of the given media. In this section, therefore, we will firstly give a brief overview of the process of MP3 compression, and then propose a 21-D feature vector to measure quantization artifacts at different compression rates.

2.1. MP3 Audio Compression

In the MP3 audio compression [10], the audio signal is firstly divided into frames of size 1152 samples with half overlapping. And then each frame is then fed to the MP3 encoder as shown in Fig. 1.

The frame is firstly separated into 32 subbands with the analysis filterbank. And then 18 frequency coefficients can be obtained after the modified discrete cosine transform (MDCT) performed on each subband. Finally we can obtain 576 frequency coefficients for each frame. Based on the properties of HAS (Human Auditory System), the psychoacoustic model is then used to analyze the resulting MDCT coefficients and get the masking thresholds which are used in the succedent quantization operation. In order to obtain a trade-off between the bit rate and quality distortions of the compressed audio, the quantization is necessary to remove some of the less audible components, and thus the quantization artifacts will be introduced in this stage. Finally, the quantized coefficients are further encoded using the lossless coding to obtain a bitstream. The decoder works in a reverse manner.

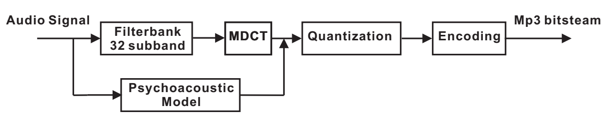

Fig. 1. The block diagram of MP3 encoder for a frame

2.2. Feature extraction

To capture the quantization artifacts, we firstly divide a questionable WAV audio into overlapping frames, which is consistent with the MP3 compression standard. And then we repeat the operations as shown in the Fig.1 and stop before the quantization operation to obtain 576 MDCT coefficients for each overlapping frame.

As illustrated in Fig. 2, two different feature sets from the MDCT coefficients will be analyzed and employed to measure the quantization artifacts of audio. Finally, these features are combined to estimate the audio compression history both for MP3 and WMA.

♦ Feature Set #1:

An obvious quantization artifact is that many MDCT coefficients will be quantized to zero after lossy compression, which means that the average number of the coefficients whose value exactly equal to zero per frame will be increased. This measure, therefore, can be used to identify whether the WAV audio has been compressed or not.

Fig. 3 shows the boxplots of the average number of zero coefficients per frame for 8,800 original uncompressed audio and their MP3 decompressed versions at the compression bit-rates of 64kbps, 96kbps and 128kbps. From Fig. 3, it is observed that the zero MDCT coefficients actually increase after compression and the features of the original uncompressed audio and their compressed versions are

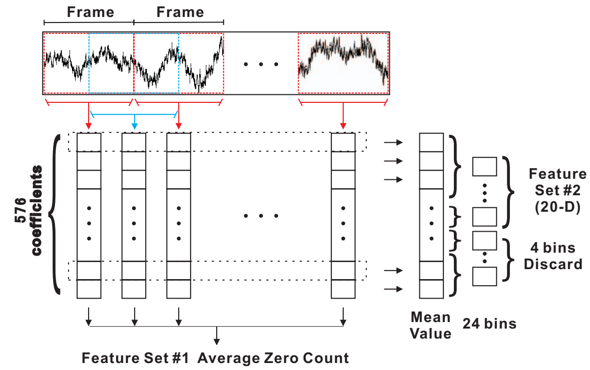

Fig. 2. Illustration of feature extraction

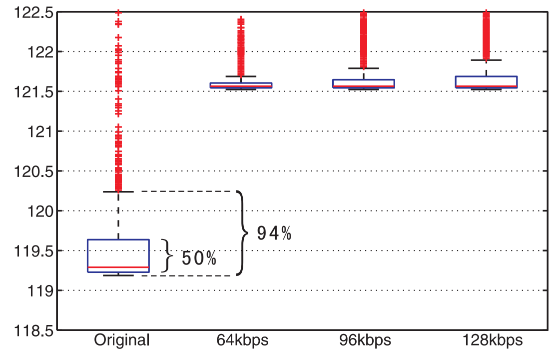

Fig. 3. The average number of zero values per frame for 8,800 original audio and their MP3 decompressed versions at different bit-rates.

mostly separated. However, for those compressed audio, the features are mainly concentrated on the range of $[121.5, 122]$. Therefore, we need other features to discriminate them.

♦ Feature Set #2:

To tell the differences among those compressed audio at different bit-rates, we investigate the MDCT coefficients of audio in different frequency components.

Please note that the 576 MDCT coefficients contain different frequency components of the corresponding frame from lower to higher frequencies, just as the 64 DCT coefficients in each 8×8 block in JPEG images along the zigzag order. It is expected that most of the higher frequency components will be weaken after compression. Usually, the severer compression we use, the more the higher frequency components will be removed, and therefore, this can serve as an useful sign for the detection of the audio compression at different bit-rates.

As shown in Fig. 2, we first calculate 576 mean values of each MDCT coefficients over all overlapping frames. Then, to reduce the dimension of the feature set, we divide the 576 mean values into 24 bins (24 coefficients in each bin) equally, and then calculate the average of the absolute value of the corresponding MDCT coefficients in each bin to obtain 24 average values. Based on our extensive experiments, we observed that the last 4 values (the highest frequency components) are usually zero even for original uncompressed audio. Therefore, we just employ the first 20 values as feature set #2.

Fig. 4 illustrates the average distributions of the 20-D features

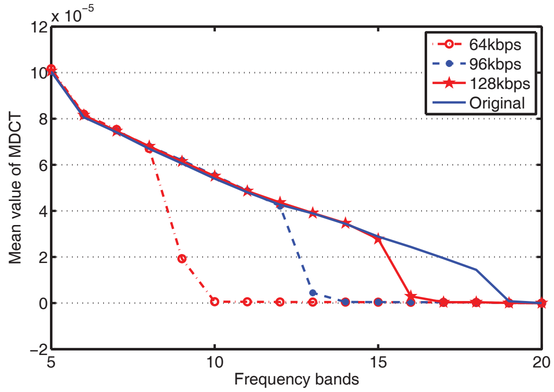

Fig. 4. The 20-D features for original audio and their MP3 decompressed versions at different compression bit-rates.

over 8,800 original uncompressed audio and their MP3 decompressed versions at the compression bit-rates of 64kbps, 96kbps and 128kbps. It is observed that the experimental results fit our analysis very well: more high frequency components will be removed with decreasing the bit-rates (severer compression). For example, when the bit-rate is 128 kbps, those average values whose bins are larger than 16 will decrease to around zero. And when the bit-rate becomes 64kbps, those average values whose bins are larger than 10 will decrease to around zero.

♦ Feature Combination:

Finally, we will combine the feature set #1 and feature set #2 to form a 21-D feature vector, which is then used to estimate the compression history of the given digital audio signal.

3. EXPERIMENTAL RESULTS

In the experiments, we collect 8,800 audio clips of 5 seconds, stereo, 44.1kHz (around 400 overlapping frames), which are randomly cut from original uncompressed WAV audio with different contents, including blues, country, disco, jazz, rock, pop and classical music. For each audio clip, we employ the Goldwave software [11] to obtain the compressed data (MP3 and WMA) with different compression rates ranging from 32 kbps to 128 kpbs. These compressed audio are then decompressed and finally stored in WAV format to remove all previous compression information.

In order to show the effectiveness of the proposed features, the following four experiments have been conducted. Feature vector is extracted as mentioned in the Section 2, and the support vector machine [12] is used for classification. In each experiment, 30% of test data is randomly selected in the training stage, and the remaining 70% is used for testing.

3.1. Results of MP3 Audio

♦ Results for a Fixed Compression Bit-rate: In this experiment, we aim to determine whether a given WAV file is an original uncompressed audio or MP3 decompressed audio with a fixed compression rate. Here, the test compression rates are 32kbps, 48kbps, 64kbps, 80kbps, 96kbps and 128kbps.

The experimental results are shown in Table 1. It is observed that the proposed features can achieve much better performance when compared with the method [3]. Even the bit-rate is as high as 128kbps, our accuracy remains over 98.24%, while it drop to 73.50% for the method [3].

Table 1. Accuracy of the classification of the original WAV file and MP3 decompressed WAV file (%).

	32kbps	48kbps	64kbps	80kbps	96kbps	128kbps
Our Method	99.65	99.87	99.81	99.69	99.73	98.24
Method [3]	93.92	89.96	89.22	86.29	82.19	73.50

♦ Results for a Random Compression Bit-rate: In this experiment, we aim to determine whether a given WAV file is an original uncompressed audio or the MP3 decompressed audio with a random compression rate. Here, the bit-rate is randomly selected from 32kbps, 48kbps, 64kbps, 80kbps, 96kbps and 128kbps. The experimental results show that our detection accuracy is 98.46%, while only 80.71% for the method [3].

♦ Confusion Matrix for Different Compression Bit-rates: In this experiment, the test data includes 8,800 original uncompressed audio and the corresponding MP3 decompressed audio with 6 different bit-rates. In all, there are 61,600 ($8,800 \times 7$) audio clips. And we aim to estimate the bit-rate of a WAV audio which is previously MP3 compressed.

The confusion matrix is shown in Table 2. The first row shows that 98.19 percent of the original WAV files are recognized as original WAV file. And the table also suggests that most audio of different bit-rates can be correctly classification.

Table 2. Confusion matrix for identifying MP3 audio at different bit-rates (%), the symbol * denotes the value less than 2.5%.

	WAV	32	48	64	80	96	128
WAV	**98.19**	*	*	*	*	*	*
32	*	**99.25**	*	*	*	*	*
48	*	2.91	**96.68**	*	*	*	*
64	*	*	*	**96.55**	*	*	*
80	*	*	*	2.60	**95.35**	*	*
96	*	*	*	*	*	**96.97**	*
128	*	*	*	*	*	*	**95.15**

3.2. Results of WMA Audio

In this experiment, we apply our features to identify the compression history of WMA decompressed audio. Three similar experiments as described in the previous subsection have been done. The experimental results are shown as follows.

♦ Results for a Fixed Compression Bit-rate: As shown in Table 3, the experimental results also show that the proposed method is effective for the WMA audio even the bit-rate as high as as 128, and it outperforms the method [3] significantly.

♦ Results for a Random Compression Bit-rate: The experimental results show that our detection accuracy is 90.90%, while 75.8% in the method [3].

♦ Confusion Matrix for Different Compression Bit-rates: The experimental results in Table 4 have shown that the proposed features can still obtain satisfactory accuracy for the WMA decompressed audio.

3.3. Results of the Mix of WAV, MP3 & WMA

For each original uncompressed WAV audio in this experiment, we obtain a MP3 and WMA decompressed audio at a random bit-rate,

Table 3. Accuracy of the classification of the original WAV file and WMA decompressed WAV file (%)

	32kbps	48kbps	64kbps	80kbps	96kbps	128kbps
Our Method	98.07	97.89	98.11	97.73	97.28	96.44
Method [3]	85.00	79.08	77.33	75.17	64.96	58.39

Table 4. Confusion matrix for identifying WMA audio at different bit-rates (%)

	WAV	32	48	64	80	96	128
WAV	**94.01**	*	*	*	*	*	*
32	*	**97.63**	*	*	*	*	*
48	*	*	**90.81**	5.77	*	*	*
64	*	*	14.93	**82.27**	*	*	*
80	*	*	8.00	3.96	**84.10**	*	*
96	3.39	*	3.39	*	*	**85.47**	4.35
128	3.57	*	*	*	*	9.25	**84.89**

respectively. We aim to identify whether a given WAV audio has been compressed with MP3 or WMA previously. As shown in Table 5, it is observed that the detection accuracies are still satisfactory for differentiating the three types of audio.

Table 5. Detection accuracy for the Mix of WAV, MP3 & WMA audio (%)

	WAV	MP3	WMA
WAV	**90.27**	*	9.16
MP3	*	**89.51**	9.89
WMA	7.53	7.20	**85.28**

3.4. Frame Offset Problem

In previous experiments, all test audio have the same frame structure. In this subsection, we try to test the robustness of the proposed features for audio with frame offset [3]. In this experiment, some samples of the test MP3 and WMA decompressed audio will be randomly cut out. Here, the number of the deleted samples is randomly from 1 to 22050 (half of a second). Two experiments have been done.

⧫ Identifying MP3 at random bitrates from original WAV: Detection accuracy is 93.80%. (98.46% for no frame offset, refer to previous experiment in Section 3.1)

⧫ Identifying WMA at random bitrates from original WAV: Detection accuracy is 89.32%. (90.90% for no frame offset, refer to previous experiments in Section 3.2)

The above results have shown that our method is still effective for those audio with frame offset.

4. CONCLUSION

Compression history identification for digital audio signal is an important issue in digital forensics. In this paper, we firstly analyze two novel feature sets from the MDCT coefficients, and then proposed a method to identify the original WAV from those WAV files that has been previously compressed by MP3 or WMA, and further to estimate the hidden compression rates. The extensive experimental results have shown the effectiveness of our method. In our future works, we will extend our features for other forensic problems, such as detection of MP3/WMA recompression and audio splicing.

5. REFERENCES

[1] C. Grigoras, "Digital audio recording analysis: The electric network frequency (ENF) criterion," The International Journal of Speech Language and the Law, vol. 12, no. 1, pp. 63–76, 2005.

[2] H Farid, "Detecting digital forgeries using bispectral analysis," Report in MIT AI Memo AIM-1657, MIT, 1999.

[3] R. Yang, Z. Qu, and J. Huang, "Detecting digital audio forgeries by checking frame offsets," in Proc. of the 10th ACM workshop on Multimedia and security, Oxford, United Kingdom, 2008, pp. 21–26.

[4] C. Kraetzer, A. Oermann, J. Dittmann, and A. Lang, "Digital audio forensics: A first practical evaluation on microphone and environment classification," in Proc. of the 9th workshop on Multimedia and security, Dallas, Texas, USA, 2007, pp. 63–74.

[5] R. Buchholz, C. Kraetzer, and J. Dittmann, "Microphone classification using fourier coefficients," in Proc. of the 11th International Workshop on Information Hiding, Darmstadt, Germany, June 2009.

[6] R. Yang and J. Huang Y. Shi, "Defeating fake-quality MP3," in Proc. of the 11th ACM workshop on Multimedia and security, Princeton, New Jersey, USA, 2009, pp. 117–124.

[7] M. Qiao, A. H. Sung, and Q. Liu, "Revealing real quality of double compressed MP3 audio," in Proc. of the international conference on Multimedia, Firenze, Italy, 2010, MM '10, pp. 1011–1014.

[8] X. Feng and G. Doërr, "FLD-based detection of re-compressed speech signals," in Proc. of the 12th ACM workshop on Multimedia and security, Roma, Italy, 2010, pp. 43–48.

[9] W. Luo, J. Huang, and G. Qiu, "JPEG error analysis and its applications to digital image forensics," IEEE Transactions on Information Forensics and Security, vol. 5, no. 3, pp. 480–491, Sept. 2010.

[10] "Information technology - coding of moving pictures and associated audio for digital storage media up to about 1.5 mbit/s," ISO/IEC International Standard IS 11172-3.

[11] "Goldwave software," http://www.goldwave.ca/.

[12] C.-C. Chang and C.-J. Lin, "LIBSVM : a library for support vector machines," ACM Transactions on Intelligent Systems and Technology, vol. 2, pp. 27:1–27:27, 2011.