EXPOSING THE DOUBLE COMPRESSION IN MP3 AUDIO BY FREQUENCY VIBRATION

Tianzhuo Wang, Xiangwei Kong, Yanqing Guo, Bo Wang

School of Information and Communication Engineering
Dalian University of Technology, Dalian, 116024, China

ABSTRACT

A novel approach is presented to detect double compressed MP3 audio by frequency vibration. With the analysis of double compression effect on MDCT (Modified Discrete Cosine Transform) coefficients in MP3 audio, we propose a simple feature called FVV (frequency vibration value) to measure the vibration caused by double compression. The experimental results on challenging dataset show that our method outperforms most of the existing methods in double MP3 compression detection, especially with a second bit-rate higher than the first one. Besides, we can also estimate the original bit-rate for a double compressed MP3 by this technique.

Index Terms—MP3, audio forensics, double compression detection, frequency vibration, MDCT coefficients

1. INTRODUCTION

The past decades witness the fast development of digital and network technologies, which had been integrated into people's life, and become an indispensable part. People are able to record the information in everyday life with the increasing popularity of mobile terminals. And everyone could share the convenience gifted by digital multimedia. However, the authenticity of digital multimedia is difficult to be guaranteed with the occurrence of intelligent editing software, and people can easily tamper the digital multimedia without professional training.

As a part of multimedia, MP3 audio is one of the most commonly used audio format, and some methods of authenticity detection [1][2][3][4] have been proposed in recent years. However, it also faces some challenges and some problems cannot be solved perfectly. Double compression is used frequently in the process of tampering in MP3 audios, as the audio must be saved after modifying. Double compression detection is then put forward, which can assign two kinds of issues [5] : Up-transcoding (corresponding to the second compression using a higher quality than the first one) and down-transcode (corresponding to the second compression using a lower quality than the first one) .

Due to the complexity of the MP3 coding process, it is very difficult to detect double compression of audios in MP3 formats. Yang et al. [6] used the number of small-value MDCT coefficients as features to detect the audio authenticity. It is based on the observation that the coefficient distribution is variant between the fake-quality MP3 and a normal one. Still, in [7], Yang et al. extracted the first 9 bins of quantized MDCT coefficients as features, and put these features into SVM classifier. In particular, global distribution, as well as band distribution is also exploited to distinguish the normal MP3 from a double compressed one. Though this method has an impressive result in detecting an up-transcoded audio, it has a limited performance in detecting a down-transcoded one. M. Qiao et al. extracted statistical features of quantized MDCT coefficients of MP3 audio, and exploited these features in a discriminative model [8][9][5]. However, these methods based statistical features are limited in application as their high computation complexity. Bianchi et al. [10] proposed a method of double compressed detection considering simulated normal MP3 file, which is achieved by removing a given number of PCM samples of decompressed audio file and recompressing the remaining samples to the same compression quality of the file. However this method still needs other statistical features to work together, which is time-consuming.

In this paper, we present a novel approach with sincerely respect to detecting double compression for both up-transcoded and down-transcoded MP3 audios, and estimate the original bit-rate for the first compression. As we know, the key issue for double MP3 compression detection is to develop effective features which are discriminative between both kinds of audios. Previous methods focus on the properties of MDCT coefficients in spatial domain, but we find that these coefficients in transform domain have better and more regular distribution. We extract a feature vector called FVV (frequency vibration value) to measure the vibration caused by double compression, and give a more effective detection. Experimental results show that our method outperforms the state of-the-art method in both up-transcoding and down-transcoding cases.

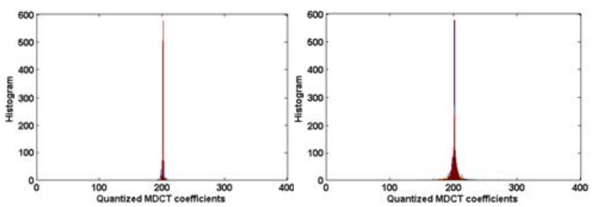
The rest of the paper is organized as follows. The detailed description of the proposed algorithm is shown in Section 2; experiments and discussion are presented in Section 3, followed by conclusion in Section 4.

2. PROPOSED ALGORITHM

2.1 Analysis of double MP3 compression

As we know, double MP3 compression contains double quantization on the MDCT coefficients, which makes the MDCT coefficients different from that of the normal one. The coefficient of a normal MP3 approximately obeys a Laplace distribution [11]. After the double compression process, this distribution can't be ensured, and this property make it possible to distinguish a normal MP3 audio from a double compressed one.

It is not easy to tell whether a coefficient distribution obey a Laplace distribution of a normal map3 audio. Figure 1 shows the coefficient distribution of a normal MP3 and a double compressed one. The figure demonstrates that the coefficient distributions of both audios are nearly the same, though with some minor differences in local.



(a) normal MP3 **(b) double compressed MP3**

Figure 1 MDCT coefficient distribution

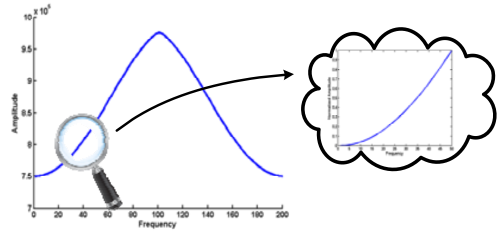
In view of this case, we consider to describe the coefficient distribution in a transform domain, for the reason that an orthogonal transform (e.g. DFT) can re-distributed the energy, and some good properties occur. We obtain two kinds of the quantized MDCT coefficients which are respectively a normal and a double compressed one with Hoffman decoding. Then MDCT coefficient distribution is computed, which is then transformed in a DFT method. We show the coefficient distribution in transform domain in Figure 2 to find out the answer whether they are still discriminatory if looking in transform domain.

In Figure 2, the normal MP3 audio has a smooth distribution in side lobe, contrary to the distribution in a double compressed MP3 audio with apparent vibration. So it is possible to detect double compression utilizing the volatility of the MDCT coefficient in transform domain.

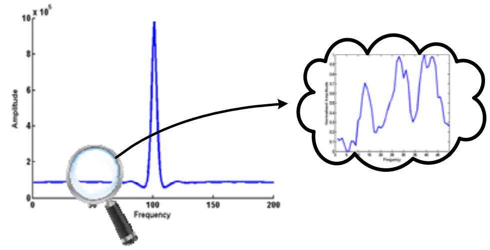
2.2 Calculation of frequency vibration value

After the observation and analysis, we certify the proposed method through mathematic derivation, and then calculate the feature vector called FVV (frequency vibration value) to measure the vibration. In general, a typical continuous Laplace distribution can be defined by

$$x_a(t) = \frac{1}{2b}\exp(-\frac{t}{b}) \qquad (1)$$



(a). The normal MP3 audio



(b). Double compressed MP3

Figure 2 MDCT distribution in transform domain

As the quantized MDCT coefficients has discrete value, its discrete Laplace distribution can be rewritten as

$$x_a(n) = \frac{1}{2b}\exp(-\frac{|n|}{b}) \qquad (2)$$

n is the quantization threshold between -200 to 200.

The causal sequence of $x_a(n)$ can be denoted as:

$$x(n) = x_a(n - N/2) = \begin{cases} \frac{1}{2b}\exp(\frac{-N/2+n}{b}) & n \le N/2 \\ \frac{1}{2b}\exp(\frac{N/2+n}{b}) & n > N/2 \end{cases} \qquad (3)$$

Then the DFT of it is as follows:

$$\begin{aligned} X(k) &= \sum_{n=0}^{N-1} x(n)W^{nk} = \sum_{n=0}^{N-1} x_a(n-N/2)W^{nk} \\ &= \sum_{-N/2}^{N/2} x_a(n)\exp(-\frac{2\pi k(n+N/2)}{N}) \\ &= (-1)^k \sum_{-N/2}^{N/2} x_a(n)\exp(-\frac{2\pi nk}{N}) \\ &= (-1)^k \sum_{-N/2}^{N/2} \frac{1}{2b}\exp(-\frac{n}{b})\exp(-\frac{2\pi nk}{N}) \end{aligned} \qquad (4)$$

Here we denote $g = \exp(-1/b)$, and Eq. (4) can be rewritten as:

$$\begin{aligned} |X(k)| &= \left|(-1)^k \sum_{-N/2}^{N/2} \frac{1}{2b} g^n \exp(-\frac{2\pi nk}{N})\right| \\ &= \left|\frac{1}{2b} g^{-N/2} \frac{1-g^N}{(1-g\cos\frac{2\pi k}{N}) + jg\sin(\frac{2\pi k}{N})}\right| \\ &= \frac{1-g^N}{2b}\frac{1}{\left[2-2g\cos(\frac{2\pi k}{N})\right]^{1/2}} \end{aligned} \qquad (5)$$

It is clear that $x(n)$ has a smooth amplitude-frequency curve. While, in a double compressed audio, the quantized MDCT coefficients distribution is different from Eq. (2), i.e. coefficients in high frequency are lost. In other words, quantized MDCT coefficients distribution is cut off in the edge. To simplify analysis, we use the distribution multiplied by an orthogonal window to simulate this process.

An orthogonal window can be rewritten as

$$h(t) = \begin{cases} A, & |t| < \dfrac{\tau}{2} \\ 0, & |t| \ge \dfrac{\tau}{2} \end{cases} \tag{6}$$

Discrete sequence $h(t)$ and causal discrete sequence $h_N(t)$ are donated as follows：

$$h(n) = \begin{cases} A, & |t| < \dfrac{d}{2} \\ 0, & |t| \ge \dfrac{d}{2} \end{cases} \tag{7}$$

$$h_N(n) = A \qquad N/2 - d \le n \le N/2 + d \tag{8}$$

So the distribution after cutting off is as follows

$$x_N(n) = x(n) \cdot h(n) \tag{9}$$

$$X_N(\text{k}) = X(k) * H(k) \tag{10}$$

Where * means convolution, $H(k)$ is the DFT of $h(n)$.

$$\begin{aligned} |H(k)| &= \sum_{0}^{N-1} h_N(n) W^{nk} = \left| \sum_{0}^{N-1} h(n - N/2) W^{nk} \right| \\ &= \left| \sum_{-N/2}^{N/2} (-1)^k h(n) W^{nk} \right| \\ &= \left| \frac{2\sin(\pi kd/N)}{1 - \cos\dfrac{2\pi k}{N} + j\sin\dfrac{2\pi k}{N}} \right| \\ &= \frac{\sin(\pi kd/N)}{\sin(\pi k/N)} \end{aligned} \tag{11}$$

It is obvious that distribution has vibration in side lobe. As a result, convolution $X_N(k)$ of $X(k)$ and $H(k)$ will also has apparent vibration in side lobe. In conclusion, we can directly use the volatility of the MDCT coefficient in transform domain to judge a double compressed MP3.

We propose T to donate the frequency vibration value (FVV), and it is defined by

$$T = \int y dx - \int y' dx \approx \sum_{i=1}^{m} |Y(i) - Y'(i)| \tag{12}$$

Where y is the MDCT coefficient in frequency domain, y' is a fitting line of the curve, m is the number of fitting, and T is the area formed by the y and y'. $Y(i)$ and $Y'(i)$ are the discrete form of y and y'.

2.3 algorithm pipeline

Recognizing the properties of MP3 double compression, we are going to present our proposed method based on the frequency vibration value. The diagram of the method is shown in Figure 3. The steps of algorithm implementation are as follows: (a) Decompress the MP3 audio file to get the quantized MDCT coefficients; (b) Transform it by DFT; (c) Calculate the FVV; (d) Determine double compression. Note that the parameter Th is only related to the property of normal and double compressed MP3 files. Its value can be predetermined through experiment with any fixed first bitrate and second bitrate.

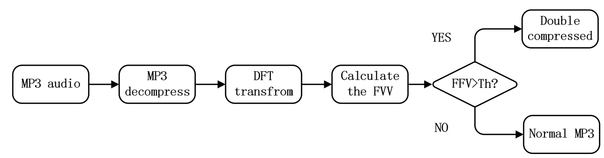


Figure 3 The diagram of the detecting method

3. EXPERIMENTAL RESULTS AND ANALYSIS

An audio dataset has been built to validate the ideas proposed in previous sections. It contains all 240 audio files, each of which in MSWAVE format and 15s long [12]. The audio files belong to four different categories: 1) speech alone: containing dialogues; 2) music alone: containing music only; 3) speech-over-music: containing speech with music; 4) environmental sound: the files relative to recording environmental. Each of WAV file has been compressed with bit-rate BR1 chosen in [64, 96, 128, 160, 192] Kbit/s. These single compressed MP3 files have been decompressed and compressed again with bit-rate BR2 chosen in [64, 96, 128, 160, 192] Kbits/s.

3.1 Effect of our proposed feature

To prove the effect of frequency vibration value, we computed the values assumed by the proposed feature, which is shown in Figure 4, where the horizontal axis represents sequence number of 100 MP3 audio files, and the vertical axis represents the feature proposed. It shows

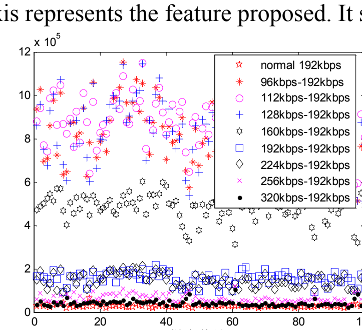


Figure 4 Comparison of the proposed feature

comparison of the proposed feature between normal 192Kbps MP3 and 8 types of recompressed MP3. 'Normal 192Kbps' means normal MP3 with the bit-rate 192Kbps.'96Kbps-192Kbps' means double compressed MP3 transcoded from 96Kbps to 192Kbps and so on The feature of normal MP3 are all near to zero, whereas those of double compressed MP3 are all lager than zero. By comparing the feature with a threshold, it is possible to discriminate these two kinds of files: double compressed files and normal ones.

3.2 Double Compression Detection

To prove the effect of proposed method based on the vibration frequency, we set up experiments using different methods. The results are shown in Tables 1,2,3,4, where BR1 means the first bitrate and BR2 means the second bitrate. The

Table 1 Detection accuracy of our method

BR2 vs BR1	64	96	128	160	192
64	100	77.7	58.3	53	53
96	100	100	77.7	74	95
128	100	100	100	100	99
160	100	100	100	100	100
192	100	100	100	100	100

Table 2 Detection accuracy of [7]

BR2 vs BR1	64	96	128	160	192
64	100	49.9	49.9	53	54
96	99.9	100	49.9	48.1	53
128	100	100	100	58	57.8
160	100	99.8	100	77	67.4
192	99.3	100	100	100	100

Table 3 Detection accuracy of [5]

BR2 vs BR1	64	96	128	160	192
64	100	53.4	64.9	64.1	59.1
96	70.9	100	63.7	67.6	64.9
128	96.7	93.2	100	65.6	70
160	98.6	99.3	93	100	75
192	99.4	99.4	95.5	82.5	100

Table 4 Detection accuracy of [10]

BR2 vs BR1	64	96	128	160	192
64	100	49.9	49.9	69.5	56.1
96	100	100	49.9	47.9	66.4
128	100	100	100	49.1	57.8
160	100	100	100	100	67.4
192	100	100	100	96.7	100

The results in table 1 show that our method is accurate and stable to detect double compressed MP3 files. In the comparison experimental, we introduce Yang et al. Method [7], Qiao et al. Method [5] and Bianchi et al. Method [10], and the results are shown in Table 2, 3, 4. Figure 5 shows detection accuracy of different method with BR1=160Kbps. With BR2 declining, all of the four methods get worse. But our method still achieves the better performance. Besides, the proposed work with only one-dimensional feature. Not only the method solves the problem of down-transcode much better, but also has higher detection accuracy in up-transcode.

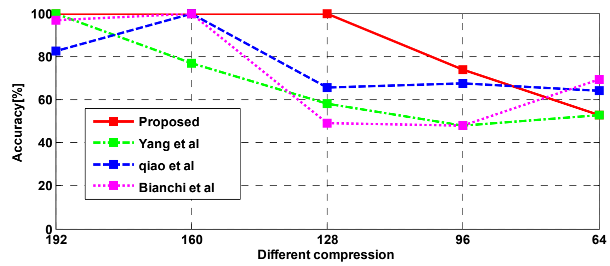


Figure 5 Detection accuracy of different method with BR1=160Kbps

3.3 Reveal the Real Bit-Rate

We utilize a Nearest Neighbor classifier to reveal the real bit-rate. Table 5 shows the confusion matrix of average multi-classification results. R means the real first compression bit-rate of double compressed MP3. E means the estimation of the first bit-rate.

Table 5. The estimate result

E vs R	64	96	128	160	192
64	99.6	0	0	0	0
96	0.4	95.8	1.4	0	0
128	0	3.2	90.8	3.5	4.4
160	0	1.0	8.8	95.5	10.6
192	0	0	0	1.0	85.0

The estimate result shows that the proposed method could reveal the real bit-rate of double compressed MP3. Similar to the results of double compression detection, the performance of revealing the real bit-rate is better in the up-transcoded than in down-transcoded.

4. CONCLUSIONS

In this paper, we present a novel and simple method to detect double MP3 compression based on the frequency vibration of MDCT coefficients. To estimate the effect of double MP3 compression on the quantized MDCT coefficients, we bring the frequency vibration value and prove its reasonability. Experimental results show that our method is very effective to detect double compression than the typical method in the field. Besides, we can also estimate the original bit-rate for a double compressed MP3. However, the detection of down-transcoding is still an issue for more people to look into it.

5. REFERENCES

[1] R. Yang, Z. Qu, and J. Huang. Detecting digital audio forgeries by checking frame offsets. In ACM Multimedia and Security Workshop, pages 21–26, 2008.

[2] Maher R. Audio forensic examination [J]. Signal Processing Magazine, 26(2): 84-94,2009.

[3] Koenig B E, Lacey D S. Forensic authentication of digital audio recordings [J]. Journal of the Audio Engineering Society, 57(9): 662-695，2009.

[4] Rodríguez D P N, Apolinário J A, Biscainho L W P. Audio authenticity: Detecting ENF discontinuity with high precision phase analysis [J]. IEEE Transactions on Information Forensics and Security, 5(3): 534-543，2010.

[5] Qiao M, Sung A H, Liu Q. Improved detection of MP3 double compression using content-independent features [C]. In ICSPCC, 2013.

[6] Yang R, Shi Y Q, Huang J. Defeating fake-quality MP3 [C]. Proceedings of the 11th ACM workshop on Multimedia and security. ACM, 117-124, 2009.

[7] Yang R, Shi Y Q, Huang J. Detecting double compression of audio signal [C]. International Society for Optics and Photonics, 2010.

[8] Liu Q, Sung A H, Qiao M. Detection of double MP3 compression [J]. Cognitive Computation, 2(4): 291-296, 2010.

[9] Qiao M, Sung A H, Liu Q. Revealing real quality of double compressed MP3 audio [C]. Proceedings of the international conference on Multimedia, 2010.

[10] Bianchi T, De Rosa A, Fontani M, et al. Detection and classification of double compressed MP3 audio tracks [C]. Proceedings of the first ACM workshop on Information hiding and multimedia security, 2013.

[11] Yu R, Lin X, Rahardja S, et al. A statistics study of the MDCT coefficient distribution for audio [C]. In ICME, 2014.

[12] http://www.ee.columbia.edu/~dpwe/sounds/musp/schei-slan.html