World Academy of Science, Engineering and Technology
International Journal of Electrical and Computer Engineering
Vol:3, No:7, 2009

A Multipurpose Audio Watermarking Algorithm Based on Vector Quantization in DCT Domain

Jixin Liu, and Zheming Lu

***Abstract*—**In this paper, a novel multipurpose audio watermarking algorithm is proposed based on Vector Quantization (VQ) in Discrete Cosine Transform (DCT) domain using the codeword labeling and index-bit constrained method. By using this algorithm, it can fulfill the requirements of both the copyright protection and content integrity authentication at the same time for the multimedia artworks. The robust watermark is embedded in the middle frequency coefficients of the DCT transform during the labeled codeword vector quantization procedure. The fragile watermark is embedded into the indices of the high frequency coefficients of the DCT transform by using the constrained index vector quantization method for the purpose of integrity authentication of the original audio signals. Both the robust and the fragile watermarks can be extracted without the original audio signals, and the simulation results show that our algorithm is effective with regard to the transparency, robustness and the authentication requirements.

***Keywords*—**Copyright Protection, Discrete Cosine Transform, Integrity Authentication, Multipurpose Audio Watermarking, Vector Quantization.

I. Introduction

THE rapid development and explosive growth of computer network and multimedia technology introduces a series of challenging problems both for the intellectual protection and the content integrity verification for the digital multimedia. Digital watermarks have been recently presented as an effective tool to provide copyright protection for high quality multimedia artworks. It is a technique to embed data into digital audio, image or video, and at the same time, introducing imperceptible distortion to the original multimedia, so that it allows ownership to be established or a legal holder to be identified. The embedded watermarks can later be detected or extracted from watermarked media for authentication or identification [1]. With decade development, digital watermarking as an effective method to accomplish copyright protection and content integrity authentication of multimedia data has been explored deeply and widely. Accordingly, the digital watermarking technique can be catalogued into robust watermarking and fragile watermarking for different goals. Generally speaking, there is only one watermark is embedded into some multimedia data which need copyright protection or content integrity authentication. But in some special circumstances, both the watermarks would need to be embedded into the same multimedia data to make the algorithm implement different purpose. For example, the owner of the multimedia artworks would hope to: use one watermark to claim ownership, use a second watermark to verify content integrity, and use a third watermark to convey a caption title [2]. However, there are only few multipurpose watermarking algorithms available, which mainly focus on the image signal in order to fulfill the application requirements [3], [4]. In [3], three blind watermarking techniques to embed watermarks in a watermark space are proposed, which are Single Watermark Embedding (SWE), Multiple Watermark Embedding (MWE) and Iterative Watermark Embedding (IWE). SWE uses two secret keys to embed a meaningful binary logo image in the watermark space, by using the spread spectrum technique and some novel features. It does not require the watermark to be orthogonal to the original data, thus allowing bit sequence embedding even in small images. Based on SWE, the second proposed technique in [3], called Multiple Watermarks Embedding (MWE), is developed to embed multiple watermarks simultaneously in the same watermark space. Different secret keys are used for different watermarks. In [4], an image multipurpose watermarking method based on the multistage vector quantization is presented, in which the semi-fragile watermark (for content authentication) and the robust watermark (for copyright protection) are embedded in different VQ stages by using different methods, and both of them can be extracted without the original image. Since the Human Auditory System (HAS) is more sensitive than Human Visual System (HVS) to the absolute phase information, multipurpose audio watermarking is more difficult to design and implement compared with multipurpose image watermarking [5]. Thus far, few multipurpose audio watermarking schemes have been presented. A cocktail watermarking scheme for audio signal is introduced in [6], where the robust and semi-fragile watermark are detected by using different detection procedures and the tampering localization is also achieved. However, its robustness against common audio signal processing manipulations is not approving. In [7], the robust watermark is embedded in the low frequency range using mean quantization, while the semi-fragile watermark is embedded in the high frequency range by quantizing single coefficient. Both the robust watermark and the semi-fragile watermark can be extracted

Jixin Liu is with Department of Electronics and Information Engineering Harbin Institute of Technology Shenzhen Graduate School, Shenzhen, China (phone: 0086-13732216189; fax: 0086-0571-86971797; e-mail: ljxok8025@ yahoo.com.cn).

Zheming Lu is with School of Aeronautics and Astronautics, Zhejiang University, Hangzhou, China (e-mail: zheminglu@zju.edu.cn).

without host audio. But the semi-fragile watermarking scheme can not achieve tampering localization. In [8], a chaotic mapping based semi-fragile watermark (for content authentication with tampering localization) is embedded in the detail wavelet coefficients based on the instantaneous mixing model of the Independent Component Analysis (ICA) scheme, while a robust watermark (for copyright protection) is embedded based on the zero-watermarking idea. In [12], a multiple watermarking scheme for speech signal is proposed by combining Linear Prediction Coding (LPC) and VQ method, the robust watermark is not really embedded in the original speech signal but extracted by using the zero-watermarking idea. And the other watermark is a semi-fragile one, which can only resist to the VQ attacks. These drawbacks restrict its application areas.

The technical challenges and requirements existing in the design of multipurpose audio watermarking schemes include [8]:

(1) How to reduce the influence of the latter embedded watermarks on the former embedded one, or how to reduce the mutual effect between the embedded watermarks?

(2) How to solve the contradiction between robustness and transparency requirements?

(3) How to guarantee the independent and blind extraction of each watermark?

(4) How to make the whole scheme resist to common audio signal processing manipulations?

According to the above requirements, a novel multipurpose audio watermarking scheme is presented in this paper. In the proposed scheme, a fragile watermark W_F (for content authentication) is embedded in the indices based on the constrained vector quantization scheme, while a robust watermark W_R (for copyright protection) is embedded based on labeled vector quantization scheme in the host audio signal A_O. Both of them are realize in the Discrete Cosine Transform domain, and the reason why we adopt this domain is that the features of Human Audio System (HAS) can be incorporated with watermarking in the transform domain more effectively, and that it is advantageous to inaudibility for the energy of embedded information in the transform domain will spread over all frequency parts. This paper is organized as follows: firstly, the introduction of the vector quantization and labeled codeword vector quantization; then the proposed multipurpose watermarking scheme is depicted detailedly, including the embedding and extraction procedure; at last, the simulation results are described and the brief discussion of the proposed algorithm is done.

II. Vector Quantization and Labeled Codeword Vector Quantization

A. Vector Quantization

VQ is an effective lossy compression method with a high compression ratio and a simple table lookup decoder. A k-dimensional vector quantizer Q of size N is a mapping from the k-dimensional Euclidean space R^k into a finite set (or codebook) $C=\{c_0,c_1,...,c_{N-1}\}$, where $c_i \in R^k$ is called a codeword and N is the codebook size. The codebook is often generated offline by the LBG algorithm [9] from a training set. The signal to be coded are first divided into vectors and then encoded by the trained codebook. For each k-dimensional input vector x, we can find its nearest codeword c_i under a certain distance metric as in the following formula:

$$d(x,c_i)=\min_{0\le j\le N-1} d(x,c_j) \tag{1}$$

Where, $d(x,c_j)$ is the distortion between the input vector x and the codeword c_j, and it can be calculated as follows:

$$d(x,c_j)=\sum_{l=0}^{k-1}(x_l-c_{jl})^2 \tag{2}$$

Then, the corresponding index i is transmitted over the channel to the decoder. The decoder holds an exact copy of the same codebook. For each index i, the decoder only performs a simple table lookup operation to obtain i and then uses it to reconstruct the input vector x. So, lossy compression is achieved by substituting a codeword for the input vector and transmitting or storing only the index of the codeword rather than the codeword itself.

B. Labeled Codeword Vector Quantization

The main idea of the proposed robust watermarking algorithm is to assign each input mid-frequency DCT-coefficient vector to the nearest codeword labeled by the corresponding watermark bit, and the proposed fragile watermarking method is to modulate the specified bit of the indices of the high-frequency DCT-coefficient vector by using the fragile watermark bits. To achieve this, two codebooks should be generated by using the LBG algorithm. The codebook for fragile watermarking is generated by using the standard LBG method. Assuming the codebook for robust watermarking has been generated by the conventional LBG algorithm, in order to recognize the watermark bits in the extraction process without the original audio signal, we assign a label being either ‘0’ or ‘1’ to each codeword. To satisfy the condition that surrounding each cluster, there should be at least a cluster labeled differently, we employ the labeling method proposed in the literature [10] as follows, which can label the codewords during the codebook generation process:

Step 1: Generate the codebook C_0 of size $N/2$ from the whole training set by using the conventional LBG algorithm.

Step 2: Partition the whole training set into $N/2$ clusters based on the nearest neighbor condition with the codebook C_0.

Step 3: Generate two codewords for each cluster, i.e. partition each cluster into two sub-clusters, by using the conventional LBG algorithm. One is labeled ‘0’, and the other is labeled ‘1’, randomly.

Step 4: Collect all the codewords to form the final labeled codebook C of size N. Record all codeword labels to form the labeling key Key_l.

After generating the labeled codebook C for the robust watermarking and the codebook C_F for fragile watermarking, we can use them to perform the watermark embedding and extracting process.

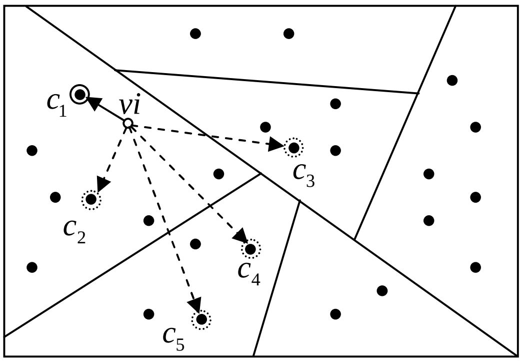

Fig. 1 The constrained index vector quantization procedure

III. Proposed Watermarking Algorithm

A. Robust Watermark Embedding Process

Step 1: The binary robust watermark W_R is first permuted by the key Key_{WR} to form a watermark sequence W_{PR} to be embedded, and w_{PR} represents all the watermark bits in W_{PR}.

Step 2: The original audio signal A_O is segmented into frames with 4-times length of the labeled codeword dimension, and the DCT transform is performed on each frame, then the mid-frequency coefficients of each frame are chosen to form a vector. All vectors compose a vector set V_R.

Step 3: For each vector $v_i \in V_R$, find its nearest codeword c_i and compare the corresponding label L_i with the watermark bit w_{PR}. If $L_i = w_{PR}$, then the codeword c_i is used to reconstruct V_i .Otherwise, select the nearest codeword c_j among all codewords labeled w_{PR} to reconstruct v_i.

Step 4: Repeat Step 3 until all watermark bits are embedded. To enhance the robustness of the watermarking algorithm, we repeatedly embed the watermark W_{PR} into the original audio signal for T times, where $T = \lfloor Length(A_o)/(FrameSize \times Length(W_{PR})) \rfloor$.

B. Fragile Watermark Embedding Process

Step 1: The binary fragile watermark W_F is first transform to 1-Dimensional to form a watermark sequence W_{PF}, and it is replicated T times according to the length of the original audio signal A_O, where T is just the same as above mentioned.

Step 2: The original audio signal A_O is segmented into frames with 4-times length of the labeled codeword dimension, and the DCT transform is performed on each frame, then the high-frequency DCT coefficients of each frame are chosen to form a vector. The dimension of it is just as the dimension of the pre-generated codebook for fragile watermark embedding, and the effect of the select of the dimension to the quality of the watermarked audio signal will be discussed also. All these vectors compose a vector set V_F.

Step 3: For each high-frequency DCT coefficient vector $v_{Fj} \in V_F$, it is quantized by the constrained index vector quantization method. The specified bit, which is selected by the key Key_{WF}, of the index is fixed to be the same as the fragile watermark bit w_{PF}. Under this condition, select the nearest codeword c_{Fj} in the fragile codebook C_F to reconstruct v_{Fj}. It can be better understood by Fig. 1. The input vector v_{Fj} should be quantized as c_1 if no watermark bits embedded, but using the constrained index vector quantization method, the input vector may be quantized as c_2, c_3, c_4, c_5 and so on. With regard to which codeword will be selected depends on the specified index bit which is determined by key Key_{WF}.

Step 4: Repeat Step 3 until all fragile watermark bits are embedded.

Step 5: Replace all the mid-frequency DCT-coefficients with the corresponding labeled codewords and the high-frequency DCT-coefficients with the corresponding fragile codewords, and then perform IDCT transform on each frame to obtain the watermarked audio signal A_w.

Fig 2 shows the block diagram of embedding the watermark W_R and W_F into the original audio signal A_o.

C. Watermark Extraction Process

Fig. 3 shows the block diagram of extracting the watermark W_R' and W_F' from the suspect audio signal A'. Segment the suspicious audio signal A' into frames with 4-times length of the codeword dimension, and perform the DCT transform on each frame, and the mid-frequency and the high-frequency coefficients are extracted and grouped to vectors. After that, these vectors are processed separately.

As for the robust watermark W_R', the extraction process is very simple and can be performed blindly only with the codeword-labeled VQ codebook C, the labeling key Key_l) and the watermark permutation key Key_{WR} .The extracting process is as follows:

Step 1: obtain the mid-frequency DCT-coefficient vector set V_R'. Establish two counting arrays: $Count0[m]$ and $Count1[m]$, where $m = Length(W_{PR})$.

Step 2: For each vector $v_i' \in V_R'$, find its nearest codeword in C and record the corresponding label according to the labeling key Key_l.

Step 3: Repeat Step 2 until the watermark is extracted for T times.

Step 4: Judge the extracted watermark bit w'_{PR} according to $Count0$ and $Count1$.

If $Count0[i] \geq Count1[i]$, Then $w'_{PR} = 0$

Else $w'_{PR} = 1$

Step 5: Repeat Step 4 for m times until all the watermark bits are obtained, then the inverse permutation is performed on W'_{PR} (composed of all extracted bits w'_{PR}) to obtain the final extracted watermark W'_R.

For the fragile watermark W'_F, we can first find the nearest codeword by the full-search method in the fragile VQ codebook C_F, and then apply bit-decomposition to its index, whose k-th bit is just the watermark bit (k is selected by the fragile watermark embedding key Key_{WF}). Finally, we assemble the bits orderly and get the extracted watermark W'_F.

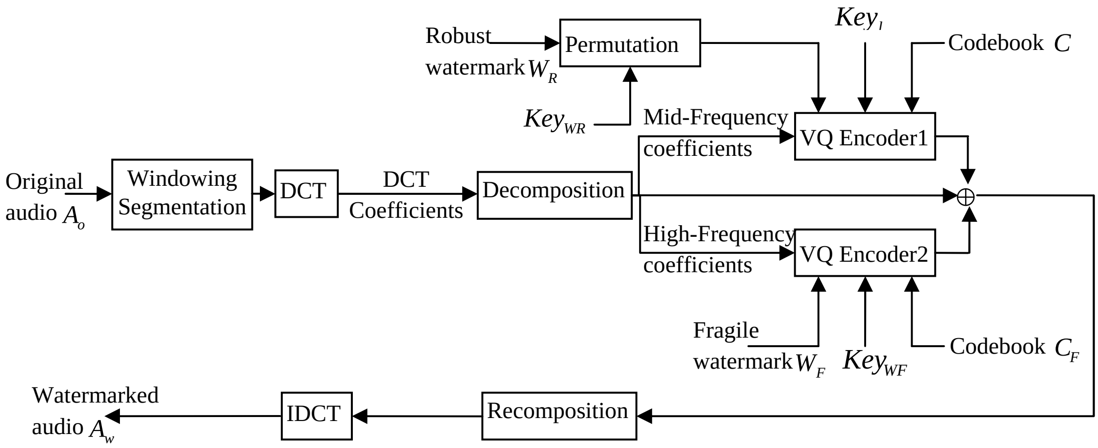

Fig. 2 The watermark embedding procedure

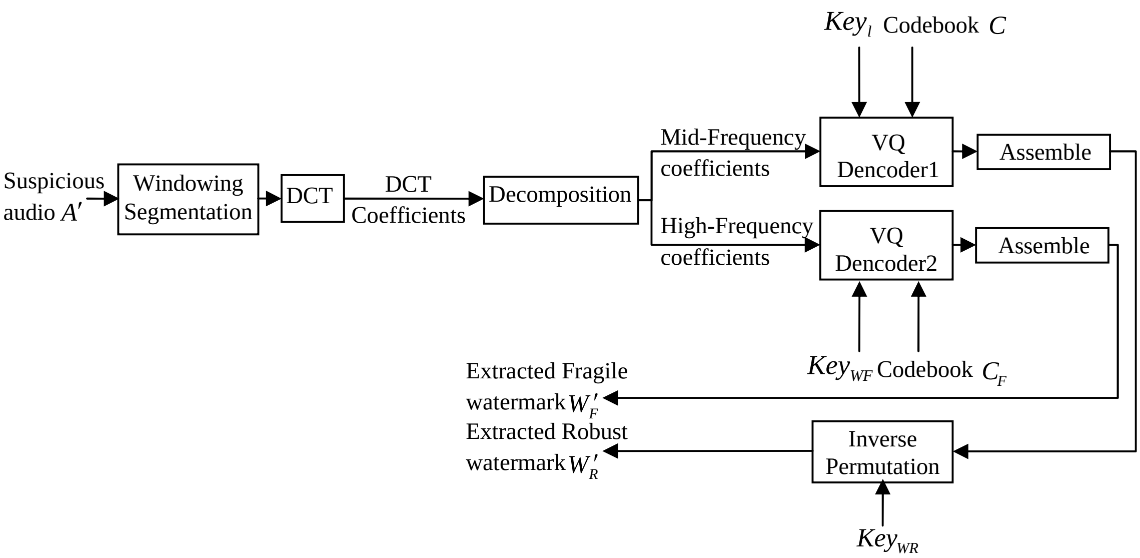

Fig. 3 The watermark extracting procedure

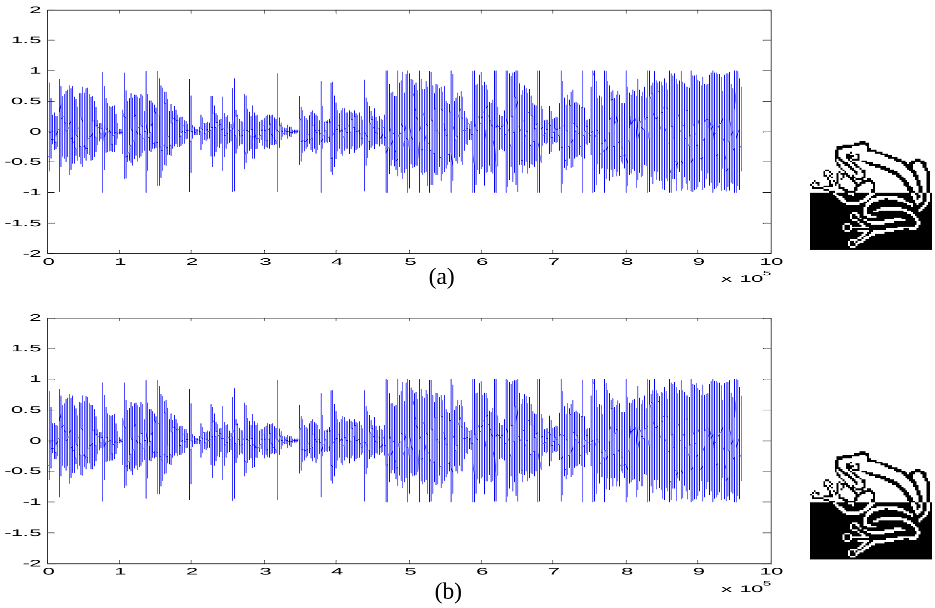

Fig. 4 (a) The original audio signal and the robust watermark,
(b) The watermarked audio signal and the extracted robust watermark without attacks.
SNR=21.87dB, NC =1.00

IV. Experimental Results and Discussions

To test the imperceptibility and robustness characteristics of the proposed audio watermarking algorithm, a set of audio signals including pop, light, rock and piano are applied. The Signal to Noise Ratio (SNR) and the subsection-averaging Signal to Noise Ratio (SNRseg) are used to evaluate the objective similarity between the original and watermarked audio signals. Normalized Correlation (NC) is used to verify the quality of the extracted robust watermark. They can be calculated by:

$$SNR = 10\log_{10}\frac{\sum_{j=1}^{m} x_j^2}{\sum_{j=1}^{m}\left|x_j - y_j\right|^2}, \quad SNR_{seg} = \frac{1}{N}\sum_{i=1}^{N} SNR_i$$

$$NC(W,W^{'}) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m} W(i,j)W^{'}(i,j)}{\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{m} W^2(i,j)}\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{m} W^{'2}(i,j)}}$$

We first test the effectiveness of our algorithm under no attacks. A clip in WAVE format (21s, mono, 16 bits/sample, 44.1 kHz) from the music titled "*I'm Gonna Be Around*" is selected, and the robust watermark is a 64×64 binary "*frog*" image, the fragile watermark is a 64*64 binary "*bird*" image. The embedding process is performed by using the labeled codebook of size 4096, each codeword having 8 dimensions, and the fragile codebook of size 1024, each codeword having 8 dimensions too. The SNR of the watermarked audio signal is 21.87dB, beyond the 20 dB requested by the IFPI. The subjective listening test shows that the watermarked audio is very similar to the original one. And the robust watermark can be extracted perfectly with NC=1.0, which is shown in Fig 4.

To evaluate the effect of different codeword dimensions and codebook sizes on the quality, a set of labeled codebooks are generated, and the watermarked audio quality with respect to the different codeword dimensions and codebook sizes are discussed. We can conclude that with the labeled codebook dimensions 4 and 8, the watermarked audio signal quality increases with the codebook size stably, and the subjective quality is very similar to the original one. Though higher SNR ratios can be obtained with higher dimensions, the subjective quality of the watermarked audio is lower and we can hear slightly annoying noises. But, to different audio signals, the quality decrease is different sharply. So we select the labeled codebook dimensions 4, 8 and 16 to evaluate the robustness of the proposed scheme. Because the embedding of the fragile watermark influences the quality of the final watermarked audio signal slightly, we just listed the result of the robustness testing as in Table I.

In the experiments for robustness evaluation, the watermarked audio signals are subjected to a variety of attacks and the NC is applied to evaluate the robustness of the algorithm. The Stirmark Benchmark for Audio v0.2 [11] and the Audacity 1.3 Beta are adopted as the editing and attacking tools in our experiments. Stirmark Benchmark for Audio is a common robustness evaluation tool for audio watermarking techniques. All the above operations listed in Table I are performed by using the default parameters in the system. From Table I, it can be found that the watermark can survive most of the common attacks.

In content authentication tests, the randomly selected watermarked audio signal samples were substituted by random samples. The detected percentage is shown in Table II. In this table, any percentage number below 100% indicates that some parts of watermarked audio have been replaced. And after finding the incorrect authentication bits, the modified content will be identified by comparing the extracted sequentially embedded fragile watermark and the original fragile watermark.

TABLE I
ROBUST PERFORMANCE OF THE PROPOSED SCHEME UNDER ATTACKS IN STIRMARK BENCHMARK FOR AUDIO V0.2

Attack Type	Robust Performance(NC)		
	Dimension4 Booksize4096	Dimension8 Booksize4096	Dimension16 Booksize 4096
AddBrumm_100	1.000	1.000	0.991
AddBrumm_9100	1.000	1.000	0.991
AddNoise_900	0.897	0.901	0.921
AddSinus	1.000	1.000	0.987
Compressor	1.000	1.000	0.932
DynNoise	0.925	0.990	0.961
FFT_RealReverse	1.000	1.000	0.992
FlippSample	0.986	0.943	0.871
LSBZero	1.000	1.000	0.982
Normalize	1.000	1.000	0.974
ExtraStereo_30	1.000	1.000	0.983
ExtraStereo_70	1.000	1.000	0.980
RC_HighPass	1.000	1.000	0.994
RC_LowPass	0.924	0.879	0.963
Smooth2	0.801	0.868	0.856
Smooth	0.813	0.859	0.831
ZeroCross	0.991	0.980	0.932
Echo	0.505	0.512	0.543
Invert	0.506	0.532	0.521
FFT_Stat1	0.622	0.598	0.743
Exchange	0.479	0.543	0.581
FFT_HLPass	0.500	0.586	0.583
FFT_Invert	0.525	0.543	0.585
CopySample	0.472	0.462	0.597
CutSample	0.584	0.567	0.432
Stat2	0.643	0.699	0.984
ZeroRemove	0.601	0.543	0.501
Amplify	0.465	0.479	0.498

TABLE II
DETECTED PERCENTAGE OF THE TAMPERED AUDIO SAMPLES

Tempered start point	Audio piece1	Audio piece2	Audio piece3	Audio piece4
200	100%	100%	98%	100%
10000	100%	95%	100%	97%
500000	97%	98%	100%	100%

V. CONCLUSION

A novel multipurpose audio watermarking algorithm is proposed based on Vector Quantization (VQ) in Discrete Cosine Transform (DCT) domain. By using this method, we can accomplish both the copyright protection and content integrity authentication at the same time. Both the robust and the fragile watermarks can be extracted blindly and the simulation results demonstrated that the presented algorithm is effective for the intellectual property protection and the integrity authentication. However, there are still some issues that are worthy of further investigation. For example, the fragile watermarking method's ability of tampering localization and distinguishing tampering types, and the psychoacoustic model can also be combined into this scheme to improve the robustness and the imperceptibility.

REFERENCES

[1] Jana Dittmann, Martin Steinebach, Thomas Kunkelmann and Ludwig Stoffels, "H204M-Watermarking for media: classification, quality evaluation, design improvements," in *Proc. ACM Multimedia 2000*, Marina Del Rey, USA, 2000, pp. 107-110.

[2] Mintzer, F., Braudaway, G.W., "If one watermark is good, are more better?" In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, USA, 1999, pp. 2067-2069.

[3] Peter H. W. Wong, Oscar C. Au,Y. M. Yeung, "A novel blind multiple watermarking technique for images," *IEEE Trans. Circuits and Systems for Video Technology*,vol.13, no. 4, pp. 813-830, August 2003.

[4] Lu, Z.M., Xu, D.G., Sun, S.H., "Multipurpose image watermarking algorithm based on multistage vector quantization," *IEEE Trans. Image Processing*, vol. 14, no. 6, pp. 822-831, June 2005.

[5] Wang, X.Y., Cui, Y.R., Yang, H.Y., Zhao, H., "A new content-based digital audio watermarking algorithm for copyright protection," In *Proc. IEEE Int. Conf. Information Security*, China, 2004. pp. 62-68.

[6] Lu, C.S., Mark Liao, H.Y., Chen, L.H., "Multipurpose audio watermarking," In *Proc. IEEE Int. Conf. Pattern Recognition*, Spain, 2000, pp. 282-285.

[7] Wang, R.D., Xu, D.W., Li, Q., "Multiple audio watermarks based on lifting wavelet transform," In *Proc. IEEE Int. Conf. Machine Learning and Cybernetics*, China, 2005, vol.4, pp. 1959-1964.

[8] Ning Chen, Jie Zhu, "Multipurpose audio watermarking algorithm," *Journal of Zhejiang University SCIENCE A*, vol. 9, no. 4, pp. 517-523.

[9] Y. Linde,A. Buzo, and R.M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Communications*, vol. 28, no.1, pp. 84-95, January 1980.

[10] Zhe-Ming LU, Wen XING, Dian-Guo XU and Sheng-He SUN, "Digital image watermarking method based on vector quantization with labeled codewords," *IEICE Trans. Information and Systems*. E86-D(12), pp. 2786-2789, 2003.

[11] Martin Steinebach, Fabien A. P. Petitcolas, Frederic Rayna, Jana Dittmann, "StirMark benchmark: audio watermarking attacks," In *Proc. IEEE Int. Conf. Information Technology: Coding and Computing*, USA, 2001, pp. 49-54.

[12] Ning Chen and Jie Zhu, "Multipurpose speech watermarking based on multistage vector quantization of linear prediction coefficients," *The Journal of China Universities of Posts and Telecommunications*, vol. 14, no. 4, pp. 64-69, December 2007.