Recoverable Video Watermark in DCT Domain

Mingzhi Cheng
College of Information & Mechanical Engineering, Beijing Institute of Graphic Communication, Beijing, China
Email: chengmz@bigc.edu.cn

Minchao Xi, Kaiguo Yuan, Chunhua Wu and Min Lei
School of Computing, Beijing University of Posts and Telecommunications, Beijing, China
Email: minchao@gmail.com, flyingdreaming@gmail.com, wchadv@gmail.com, leimin@bupt.edu.cn

Abstract—Few of current video watermarking algorithms can prevent HD videos from unauthorized copying after the production is distributed. This paper proposes a recoverable video watermarking algorithm in DCT domain to achieve this goal. Watermark is first split and modulated by CDMA. Then, after applying boundary detection, frames in one shot are divided into blocks. Finally, watermark is embedded into DCT coefficients of these blocks. The embedded watermark could resist to conventional video processing such as loss compression, resolution change, and format change, etc. The security of embedded watermark is based on the key used in CDMA modulation. With correct user keys, authorized user can completely remove the watermark and restored the high quality video. The experimental results show that the proposed algorithm is robust and practical.

Index Terms—recoverable watermark, video watermark, CDMA, DCT

I. INTRODUCTION

With the rapid growth of Internet, the publication and distribution of digital media is easier and faster. Therefore, there is an increase demand of copyright protection. Traditionally, encryption and access controlling techniques were adopted to protect the ownership of digital contents. These techniques, however, can’t prevent the media from unauthorized copying after the production is distributed. In this paper, we propose a new algorithm for video watermarking purpose.

In this paper, a new watermarking algorithm, which combines the revisable watermarking theory and video watermarking schemes, is proposed. Benefit from the feature of revisable watermark, though not perfectly restored, distortions caused by watermarking could be removed. Restoring procedure could be plugged into the video decoder, which is used when videos are playing. That means, only the authenticated video could be restored, and only authenticated users could enjoy the video of HD quality.

The rest of this paper is organized as follows. Fundamental concepts of CDMA, definition of revisable watermark and shot boundary detection are introduced in section 2. Procedures of the proposed scheme- watermark preprocess, watermark embedding and detection, and video recovery- are represented in detail in section 3. In section 4, performances of our scheme are evaluated and analyzed. Finally, conclusions and future works are raised in section 5.

II. RELATED WORK

A. Video Watermarking Techniques

Video watermarking refers to embedding watermarks in a video sequence in order to prevent the video from unauthenticated copying and identify manipulations [1]. The techniques can be divided into methods that work on compressed or uncompressed data. The watermarking techniques can be applied either in the spatial domain or in the frequency domain [2,3]. Spatial domain watermarking techniques embed the watermark by modifying the pixel values of the host video directly. They have low computational complexities and are used in video watermarking where real-time performance is concerned. But the need for absolute spatial synchronization and the weakness against frequency filtering and compression lead to disadvantages in application.

Frequency domain watermarks, especially the DCT domain watermarks, are widely used. Though higher computational complexities are required, the benefits obtained from these schemes are worthwhile. For video watermarking, a number of format-specific watermarking techniques have been proposed, including approaches based on GOP modifications, high frequency DCT coefficient manipulation, DCT block classification, etc. Considering majority of conventional video encoders use DCT coefficients as their data format, in order to be compatible with most of video encode formats, the proposed watermarking scheme works in DCT domain.

We have surveyed on the current video watermarking technologies. It is noticed that few of the current revisable watermarking algorithms can resist all conventional video attacks, including loss compression (bit rate reduction), resolution change and encode format change, etc. With this finding, we propose a new revisable video watermarking scheme based on scene detection and CDMA.

B. Revisable Watermarking

Revisable watermarking is a special kind of digital watermarking technique. It not only provides the
authentication of the copyright by embedding the predefined watermark into the original media, but also can recover the original media from the embedded one. Similar to traditional watermarking schemes, revisable watermarking schemes have to meet all requirements of the conventional watermarking such as robustness and imperceptibility [4]. Majority of revisable watermarking schemes work on spatial domain, while in this paper, we will propose a new revisable watermarking scheme applied in DCT frequency coefficients.

The revisable watermark could be classified into three types: applying data compression [5-7], using difference expansion [8,9], and using histogram bin shifting [10-12]. For the first category, the existing schemes belonged to this type are lack of robustness for the applied compression methods cannot resist the conventional distortions. The schemes using difference expansion is also weak against attacks because the location map using in these schemes is easy to be destroyed. The type of histogram bin shifting is somewhat robust but could not meet the demands of video watermarking.

Most of revisable watermarking schemes are discussed in the field of image watermarking, and most of them work in spatial domain. In the proposed scheme, revisable watermark is embedded in DCT frequency domain.

III. PROPOSED METHOD
A. Watermark Preprocess

We process watermark information with CDMA - a kind of spread spectrum technique - before embedding procedure. Pickholz et al.[13] defined spread spectrum communications as follows:

Spread spectrum is a means of transmission in which the signal occupies a bandwidth in excess of the minimum necessary to send the information; the band spread is accomplished by a code which is independent of the data, and a synchronized reception with the code which is independent of the data, and a synchronized reception with the code at the receiver is used for dispreading and subsequent data recovery.

Some authors have discussed the fundamental information theoretic limits to reliable communication to digital watermark [14-16]. The smaller is the number of bits of payload contained in a watermark, the greater is the chance of it being communicated without error.

(1) Encoding Processing

CDMA is a method for encoding binary messages. Suppose we are given a message which is in binary form of 
\[ \text{\textit{B}} = b_1, b_2, ... , b_L \] where \( b_j \in \{ 0, 1 \} \) are the bits. For symbols \( b_1, b_2, ... , b_L \), we generated \( \vec{r}_1, \vec{r}_2, ... , \vec{r}_L \) as independent zero mean pseudorandom vectors corresponding, which length is \( N \). Finally, the sequence of symbols is encoded as the summation:

\[ \vec{m}(t) = \sum_{i=1}^{L} s(b_i) \cdot \vec{r}_i(t), \quad t = 1, 2, ..., N \] (1)

Where \( s(x) = \begin{cases} 1 & \text{if } x = 1 \\ -1 & \text{if } x = 0 \end{cases} \)

Then we get a vector \( \vec{m} \) whose length is \( N \), which will be embedded into original video.

The specific choice of method and the parameter for generating the pseudorandom sequence \( \vec{r}_1, \vec{r}_2, ... , \vec{r}_L \) has direct relationship with the reliability and security of the embedded watermark. So, keys could be used as input parameters of pseudorandom sequence generator. Conventional generators include Gold Codes, Kasami codes, m-sequences, Legendre sequences and perfect maps[17-21].

An obvious limitation is recognized here: longer the original symbol sequence \( \vec{B} \) is, larger the values of elements of \( \vec{m} \) are. To avoid superabundant modification when watermark embedding, the message sequence should be limited to no longer than 320 bits (that is, for each frame, the maximum length of message is 40 bytes). So, before CDMA modulation, messages should be separated into segments in advance, each part is no longer than 320.

To distinguish different watermark segment, and to identify the validity of detected watermark in detection procedure, we should also embed a segment index identifier into the video frame together with the segment itself. The identifier could append to the message bits before CDMA modulation is applied.

(2) Decoding Processing

In most cases, vector \( \vec{m} \) will be transmitted on some carrier data, such as pixel values, frequency coefficients, etc. Therefore, without loss of generality, in decoding procedure, we can hardly get the same vector \( \vec{m} \) in encoding processing, but \( \vec{m}' \) instead. \( \vec{m}' \) could be expressed as:

\[ \vec{m}'(t) = \vec{m}(t) + \delta(t), \quad t = 1, 2, ..., N \] (2)

Where \( \delta \) stands for distortions or noise signals imposed on the original vector \( \vec{m} \).

Correspondingly, to decode the pseudorandom vector \( \vec{m}' \), vectors \( \vec{r}_1, \vec{r}_2, ... , \vec{r}_L \) are generated in turn. The decoding procedure could be expressed as formula:

\[ b_i = \begin{cases} 0 & \text{if } \vec{m}' \cdot \vec{r}_i \geq 0 \\ 1 & \text{if } \vec{m}' \cdot \vec{r}_i < 0 \end{cases} \quad i = 1, 2, 3, ..., L \] (3)

Where \( \cdot \) denotes inner product of two vectors.

Then we get the detected binary message \( \vec{B}' = b_1, b_2, ... , b_L \), where \( b_j \in \{ 0, 1 \} \).

B. Video Preprocess

Due to the temporal dimension existing in video, another problem comes to be the extended difficulty of satisfying the imperceptibility criterion, compared with
image watermarking. Because of the differences between the intensities of pixels at the same position in two adjacent frames, a scheme which embeds different watermarks into successive frames usually yields a flicker effect in video. In order to avoid obvious visual distortions in the watermarked video, we utilize video shot boundary detection.

Video shot boundary detection techniques could split video into scenes, which is consists of frames with similar contents. To avoid flicker effect, frames in one scene should contain the same watermark information. And because of the limitation of CDMA modulation mentioned above, we would embed each message segment into each corresponding scene.

Boundary detection is a component which could be dynamically replaced in our watermarking scheme. To fulfill the efficiency requirements of video watermarking, we adopt one of the conventional simple techniques such as histogram based algorithm [22, 23].

### C. Watermark Embedding

For each video scene, one corresponding watermark segment is embedded into its every frame repeatedly. So the topic turned to the embedding pattern for each frame. It is composed of 4 steps.

Step 1: Each frame is divided into the blocks of 16*16. The video signal for each block is denoted as $I_b(x, y)$, where $x$ and $y$ is the spatial coordinate in a block, $b$ is block index. For each block, the intensities are converted into 4 DCT coefficient sub-blocks of the size of 8*8. The final transformed signal is denoted as $C_{b,k}(x, y), k = 1, 2, 3, 4$ , where $k$ stands for the sub-block index.

Step 2: After applying CDMA modulation to corresponding watermark segment, we get $\tilde{m}$ for this frame. Since the proposed method should spread the watermark among all 16*16 blocks, the length of each $I_b$ must reach the amount of blocks in one frame to best effort.

Step 3: For each 16*16 block, there’s a corresponding element of $\tilde{m}$, which is an integer. In each 8*8 sub-block of this block, we choose a group of coefficients as the embedding district. From top-left to bottom-right, the 4 sub-blocks are indexed as 1~4. For these sub-blocks, we apply the following modification:

$$C_{b,k}^\prime (x, y) = C_{b,k}(x, y) + \text{strength}(x, y) \cdot \tilde{m}(b), k = 1, 2, 3, 4, b = 1, 2, 3, ..., N$$

Where $(x, y) \in SG$ , $SG$ stands for the embedding district, we choose $SG = \{(2,4),(3,3),(4,2),(2,5),(3,4),(4,3),(5,2)\}$ in the experiment of section 4. Amount of positions in embedding district is denoted as $T$.

$\text{strength}(x, y)$ is watermark strength, which could be different value for different position. We choose $\text{strength}(x, y) = 0.8$ , $(x, y) \in SG$ in experiment of section 4.

Step 4: After modifications in last step, each sub-block is converted by inverse DCT transformation.

### D. Watermark Detection

When detection procedure is applied on a suspicious video, we don’t need to detect scenes. Each frame is transmitted into this procedure to detect the information it contained. It is composed of 5 steps.

Step 1: Just like step 1 of watermark embedding, each frame is divided into blocks of 16*16. Then the intensities are converted into 4 DCT coefficient sub-blocks of the size of 8*8 for each 16*16 block. The final transformed signal is denoted as $C_{b,k}(x, y), k = 1, 2, 3, 4$ , where $k$ stands for the sub-block index.

Step 2: From top-left to bottom-right, we detect $\tilde{m}$ from 4 sub-blocks by the following formula:

$$\tilde{m}^t_i(b) = C_{b,k}^\prime [SG(t)] + C_{b,2}^\prime [SG(t)] - C_{b,3}^\prime [SG(t)]$$

$$- C_{b,4}^\prime [SG(t)] , t = 1, 2, 3, ..., T$$

Where $SG(t)$ stands for the embedding position of number $t$. For example, suppose $SG = \{(2,4),(3,3),(4,2),(2,5),(3,4),(4,3),(5,2)\}$ , then $SG(0) = (2,4), SG(1) = (3,3)$. By this step, we get $\tilde{m}$ for each position of the embedding position $SG(t)$.

Step 3: CDMA decoding process is applied on each $\tilde{m}$ in this step. For each $\tilde{m}$, a sequence of detected symbols is generated:

$$\tilde{B}_t = b_{1,t}^\prime, b_{2,t}^\prime, ..., b_{L_t}^\prime , t = 1, 2, 3, ..., T$$

© 2013 ACADEMY PUBLISHER
Step 4: All $\tilde{B}_t^i$, $t = 1, 2, 3, ..., T$, are integrated to one sequence $\tilde{B}$, using the following formula 8:

$$
\tilde{B}(k) = \begin{cases} 
0, & \text{if } \sum_{t=1}^{T} b_{k,t} \geq T / 2, \\
1, & \text{if } \sum_{t=1}^{T} b_{k,t} < T / 2
\end{cases}, \quad k = 1, 2, 3, ..., L
$$

(7)

$\tilde{B} = \tilde{b}_1, \tilde{b}_2, ..., \tilde{b}_L$ where $\tilde{b}_{i} \in \{0, 1\}$ is the final watermark segment and its corresponding index identifier we detected from this frame.

Step 5: After finish detection procedure on each frame, we’ll assemble the watermark segments according to the segment index identifier containing in $\tilde{B}$.

E. Video Recovery

Video recovery is composed of 4 steps.

Step 1: Before watermark recovery procedure, watermark detection should be applied on each frame first. Then, we get $\tilde{B} = \tilde{b}_1, \tilde{b}_2, ..., \tilde{b}_L$.

Step 2: After applying CDMA encoding on $\tilde{B}$, $\tilde{m}$ is gotten, from which we can calculate modification on each DCT coefficient when watermark embedding.

Step 3: Each frame is divided into the blocks of 16*16. For each block, the intensities are converted into 4 DCT coefficient sub-blocks of the size of 8*8. The final transformed signal is denoted as $C_{b, k}(x, y), k = 1, 2, 3, 4$.

Step 4: For each 16*16 block, there’s a corresponding element of $\tilde{m}$, which is an integer. From top-left to bottom-right, we apply the following modification on the sub-blocks, which is just the opposite of embedding:

$$
\begin{align*}
C_{b, k}(x, y) &= C_{b, k}(x, y) - strength(x, y) \ast \tilde{m}(b), k = 1, 2 \\
C_{b, k}(x, y) &= C_{b, k}(x, y) + strength(x, y) \ast \tilde{m}(b), k = 3, 4 \\
b &= 1, 2, 3, ..., N
\end{align*}
$$

(8)

IV. EXPERIMENTAL RESULTS AND ANALYSIS

The performance of proposed scheme is evaluated from two aspects: robustness and imperceptibility.

1) PSNR (Peak Signal-to-Noise Ratio) is used to evaluate the imperfections of a watermark. The definition is as follows:

$$
PSNR = 10 \log \frac{255^2}{\sum \left[ F - F^e \right]^2} 
$$

(9)

$F$ means the original frame, $F^e$ means the embedded frame. N and M stand for the width and height (resolution) of the frames. 255 mean the maximum of the pixel value.

2) The robustness could be evaluated by BER (Bit Error Rate), which is the percentage of differences between extracted watermark and embedded watermark.

$$
BER = \frac{N_d}{N} \times 100\%
$$

(10)

$N_d$ stands for the number of differences between extracted and embedded watermark stream (bit stream). N denotes the number of bits of watermark information.

The experiments are conducted based on 3 typical videos: foreman, flowers, and waterfall. Property of original videos and the video quality after embedding are listed below:

<table>
<thead>
<tr>
<th>File Name</th>
<th>Resolution</th>
<th>Frame Rate</th>
<th>Bit Rate</th>
<th>PSNR (EMBED)</th>
<th>PSNR (RESTORE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>forman_cif</td>
<td>352x288</td>
<td>25 fps</td>
<td>30.4 Mbps</td>
<td>34.1</td>
<td>54.1</td>
</tr>
<tr>
<td>flower_cif</td>
<td>352x288</td>
<td>25 fps</td>
<td>30.4 Mbps</td>
<td>34.1</td>
<td>46.2</td>
</tr>
<tr>
<td>waterfall_cif</td>
<td>352x288</td>
<td>25 fps</td>
<td>30.4 Mbps</td>
<td>34.1</td>
<td>60.4</td>
</tr>
</tbody>
</table>

Frame before and after watermark embedding are shown below:

1) Frame of forman_cif_emb.avi:

![Forman_cif_emb.avi original frame](image1)

![Forman_cif_emb.avi embedded frame](image2)

2) Frame of flower_cif_emb.avi:

![Flower_cif_emb.avi original frame](image3)

![Flower_cif_emb.avi embedded frame](image4)
We’ve evaluated the robustness of our scheme through 3 kind of video attacks: loss compression, resolution change, and format change. Attacks are conducted by ffmpeg.exe version N-34031-ge403a97.

1) Experiment With Loss Compression
Loss compression is conducted by H.264 encoder. The compression level could be measured through bit rate of the video.

<table>
<thead>
<tr>
<th>Bit Rate</th>
<th>foreman’s BER (%)</th>
<th>flowers’s BER (%)</th>
<th>waterfall’s BER (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1150kbps.avi</td>
<td>5.73</td>
<td>16.20</td>
<td>0.56</td>
</tr>
<tr>
<td>3500kbps.avi</td>
<td>0.96</td>
<td>3.12</td>
<td>0.10</td>
</tr>
<tr>
<td>8000kbps.avi</td>
<td>0.62</td>
<td>1.89</td>
<td>0.04</td>
</tr>
</tbody>
</table>

Average BER for 3 bit rate attack levels, the capacity of watermark is 40 bits.

2) Experiment With Resolution Change
Resolution change is conducted by ffmpeg.exe version N-34031-ge403a97.

<table>
<thead>
<tr>
<th>Resolution</th>
<th>foreman’s BER (%)</th>
<th>flowers’s BER (%)</th>
<th>waterfall’s BER (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>176x144</td>
<td>18.90</td>
<td>21.10</td>
<td>11.30</td>
</tr>
<tr>
<td>212x172</td>
<td>8.42</td>
<td>12.10</td>
<td>6.16</td>
</tr>
</tbody>
</table>

Average BER for 5 resolution attack levels, the capacity of watermark is 40 bits.

3) Experiment With Format Change
Format change is conducted by ffmpeg.exe version N-34031-ge403a97.

<table>
<thead>
<tr>
<th>Format</th>
<th>foreman’s BER (%)</th>
<th>flowers’s BER (%)</th>
<th>waterfall’s BER (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>H.264</td>
<td>0.59</td>
<td>1.70</td>
<td>0.04</td>
</tr>
<tr>
<td>MPEG4</td>
<td>2.73</td>
<td>8.33</td>
<td>0.35</td>
</tr>
</tbody>
</table>

Average BER for 2 kind of video encode standards, the capacity of watermark is 40 bits.

H.264 and MPEG4 are two kind of video standards that widely used in multi-media publication.

The results of evaluations demonstrate that the proposed algorithm is robust to loss compression, resolution change, and format change attack. The performance of proposed scheme turned to have the following features: video with higher texture complexity could bear severer attacks. For example, the video of waterfall has lower BER than flower’s video whose pictures have more flat areas. More works are needed with the performances of resisting resolution change attack, and the performances of different embedding positions could have more evaluations.

V. CONCLUSION AND FUTURE WORK
A revisable video watermark in DCT domain is proposed based on CDMA. The process of this watermarking scheme, including watermark preprocessing, video preprocessing, watermark embedding, detection, and video recovery, is described in detail. The experiments are conducted to demonstrate that our method can survive typical video attacks, such as loss compression, resolution change, and format change.

This proposed watermarking scheme can further be associated with different applications. Restoring procedure could be plugged into the video decoder, which is used when videos are playing. So, only the authenticated video could be restored, and only authenticated users could enjoy the video of HD quality.

ACKNOWLEDGMENT
This paper is supported by Beijing Natural Science Foundation under Grant No. 4122026 and Scientific Research Common Program of Beijing Municipal Commission of Education under Grant No. KM201210015007, KM201210015006, KZ201210015015.

REFERENCES


Mingzhi Cheng was born 1974 in Hubei China. He received the B.S. degree in Communication Engineering from Chongqing University of Posts and Telecommunications in 1996 and the M.S. degree in 2002 and the Ph.D. degree in Cryptography from Beijing University of Posts and Telecommunications in 2010.

From 2010, he worked in Beijing Institute of Graphic Communication as a lecturer. Dr. Cheng is working in packaging anti-counterfeiting and Digital Content Security.

Minchao Xi was born 1987 in Shanxi China. He received the B.S degree in singal and information processing from Beijing University of Posts and Telecommunications in 2009.

Dr. Yan is working in information security as a lecturer in Beijing University of Posts and Telecommunications.

Kaiguo Yuan was born in 1982 in Guizhou China. He received the Ph.D degree in computer science and technology from Shanxi University in 2009, and will receive the M.E. degree in computer science and technology from Beijing University of Posts and Telecommunications (BUPT) in 2013.

Xi Minchao is working in Image and Video Watermarking Theory as a graduate student in BUPT.

Chunhua Wu was born in 1977 in Hubei China. She received the B.S degree in communication engineering from Chengdu University of Information Technology in 1999 and the Ph.D degree in singal and information processing from Beijing University of Posts and Telecommunications in 2008.

Dr. Wu is working in information security as a lecturer in Beijing University of Posts and Telecommunications.

Min Lei was born 1979 in Jiangxi China. He received the B.S. degree of Science in Computer Science from Nanchang University in 1999, the M.S. degree in Computer software and theory from Beijing University of Posts and Telecommunications in 2002 and Ph.D degree in Cryptography from Beijing University of Posts and Telecommunications in 2011.

Dr. Lei is working on information security as a lecturer in Beijing University of Posts and Telecommunications.