Rate Controlled Video Watermarking
Based on Optimal DEM Codewords

Xiaoyan Liu, Lihong Ma, Hanqing Lu, Kang Cai
School of Electronic Engineering, South China University of Technology
National Lab. of Pattern Recognition, Inst. of Automation, Chinese Academy Of Sciences
Guangzhou, China, 510641, Beijing, China 100080
xyliu1982@163.com, eelhma@scut.edu.cn, luhq@nlpr.ia.ac.cn

Abstract—This paper addresses the rate maintenance requirement of video watermarking and proposes a differential energy watermarking (DEM) scheme. We embed a watermark bit by shifting the energy difference between horizontal edge components and the vertical ones in a quantized luminance block within a quantization cell, and spread the slight shift over coefficients for good quality and robustness. Since a slightly changed coefficient will possibly share the same segment code and a similar code-length with its un-watermarked precedent, our method could simplify the rate allocation strategy. Experimental results demonstrate that our DEM scheme performs better on visual quality impact, capacity, robustness and rate maintenance than the DEW (Differential Energy watermarking) scheme.

I. INTRODUCTION

Nowadays, more and more video watermarking algorithms are proposed to protect the property right of video sequence. For watermarking, robustness and capacity are very important requirements. Besides, the bit rate changes caused by watermark embedding should be considered. Langelaar et al. proposed a Bit Domain Labeling Algorithm to avoid buffer overflow or empty, and minimize the MSE for a given overall encoding rate to make the bit stream keep its original size[1]. They embedded a label signal to the least significant bits of selecting code-words, each of them existing another VLC of the same code length, with the same zero-run length and only 1 level difference of the code. Though this scheme is less computationally expensive, it is not robust to attacks of watermark removal. Another watermark scheme Differential Energy Watermarking(DEW) [1,2] embedded watermark bits by enforcing the energy differences of label-carrying region (lc-region) by selectively discarding some high frequency DCT coefficients of the region. DEW could resistant a wide range of attacks like transcoding and re-labeling, but it also has limitations. In DEW scheme, the optimal number of blocks for 1-bit labeling and the minimal cutoff point for coefficient discarding should be carefully selected. The number of DCT blocks of a lc-region for 1 bit embedding influences its watermarking capacity and robustness. The more blocks are used to embed a bit, the more robust the watermark would be, but the less capacity is achieved. On the other hand, the minimal cutoff point determines the number of discarded coefficients for enforcing the energy difference to a desired value. If too many frequency components are removed, the artifacts will be visible, and the total codeword length decreases greatly, that is, bit rate change occurs. Another typical method of compressed video watermarking was emphasized to bit rate control[3]. It embedded a label bit in a nonzero coefficient only when the host codeword was longer than its alternate watermarked codeword. Drift compensation was also used to balance the changed bit amount accumulated from pervious frames. However, its complex compensation caused problems of embedding positions convey and rate re-assignment in decoding.

To overcome the above drawbacks, we propose a DEM watermarking algorithm which is suitable for rate controlled watermarking of video streams. In this approach, dither modulation (DM) to different energy (DE) is applied to embed a bit by quantizing the energy difference between two sub-regions of a block to the nearest dither value. Since its embedding error could spread over coefficients of a lc-region, these coefficients are just slightly changed, and have strong possibility of sharing with their un-watermarked precedent the same segment code and the similar codeword length. Thus, DEM can easily meet the requirement of rate maintenance. Simulation results demonstrate that:
(1) Compared to the bit rate control scheme[3], DEM scheme is quality fidelity and needn’t convey embedded positions;
(2) Compared to DEW, our scheme is superior in rate control, large capacity and visual quality preserving.

II. ENCODING OF AC COEFFICIENTS AND DIFFERENTIAL ENERGY WATERMARKING

A. Encoding of AC coefficients
For high compression efficiency and simple algorithm, Variable Length Coding (VLC) is widely used for data compression. In many image/video coding standards, VLC is usually employed with Run Length Coding (RLC) to reduce the data rate, by combing successive zeros and a succeeded non-zero value as one symbol which can be coded in a more compact codeword [4].

In coding process, a quantized DCT block is firstly rearranged as a sequence in a zig-zag (zz) order. Since many coefficients are zero, runs of zeros are firstly identified. Each non-zero AC coefficient in a zig-zag scanned sequence is described by a composite 8-bit value [5], in the form of

\[
\text{RS=binary 'RRRRSSSS'}
\]

where 'SSSS' are 4 least significant bits, defining a category for the amplitude of a non-zero coefficient. The 4 most significant bits 'RRRR' represent the zero run-length between two non-zero neighboring coefficients. The composite value 'RRRRSSSS' is Huffman coded followed by an additional sign bit. TABLE I defines the magnitude ranges assigned to each value of SSSS.

In our scheme, we keep the ‘RRRR’ constant and slightly change 10 non-zero coefficients in a lc-region to embed one watermark bit. The total variation of 10 changed coefficients is less than a quantized step size \(\Delta\) (usually \(\Delta = 3, 4, 5\)). Therefore, after embedding a watermark bit, no more than 3–5 coefficients have to be changed within 1 level, i.e. the variation of SSSS ≤ 1, and the RRRR could keep constant, then the maximum variation of the total code length can be easily adjusted to not more than 2. Thus DEM algorithm can satisfy the requirement of rate maintenance better.

### B. Differential Energy Labeling

The DEW algorithm presented a watermark bit by energy difference (DE) of a lc-region, which was usually composed of groups of \(8 \times 8\) DCT blocks in I-frames [1, 2]. The sign of DE was enforced to a desired value by selectively discarded certain high frequency components of the lc-region. The number of \(8 \times 8\) blocks in a lc-region influenced the watermark robustness and visual impression. The larger the DE was, the more robust the watermark would be, but the more DCT coefficients were discarded, and the worse visual quality would be.

#### Step_1 Differential Energy Calculation

Firstly, computed energy \(E_c\) of the top half lc-region \(A\) and \(E_b\) of the bottom half \(B\) with the high frequency DCT coefficients (shown as the white triangular areas in Fig.1, denoted as subset \(S(c)\)). The subset \(S(c)\) was defined as:

\[
S(c) = \{ u \in [0, ..., 63] \mid (u > c) \}
\]

where \(u\) represented coefficient locations after a cutoff point \(c\) in the zig-zag scanning, \(c\) should be bigger than the minimal cutoff index \(C_{min}\) decided by optimization. Energy \(E_c\) was the squared sum of \(S(c)\) in sub-region \(A\).

\[
E_c = \sum_{b=0}^{n-1} \sum_{u \in S(c)} (\text{DCTcoeff}_{u,b})^2
\]

Where \(b\) denoted the number of blocks of sub-region \(A\). Energy \(E_b\) of sub-region \(B\) could be defined similarly.

Secondly, differential energy \(D\) between sub-region \(A\) and \(B\) was obtained by

\[
D = E_c - E_b
\]

The value of \(D\) influenced the watermark robustness and visual impression. The larger the \(D\) was, the more robust the watermark would be, but the more DCT coefficients were discarded, and the worse visual quality would be.

#### Step_2 Label Bit Embedding

Firstly, employed the sign of differential energy \(D\) to represent a label bit. If \(D > 0\), label bit “0” was represented. If \(D < 0\), label bit “1” was defined.

Secondly, adapted \(E_c\) or \(E_b\) to manipulate the energy difference \(D\) to embed a label bit:

If a label bit was “0”, all DCT components after the cutoff point \(c\) in each block of region \(B\) were eliminated by forwarding the EOB marker of each block, thus \(E_b\) was set to zero to make

\[
D = E_c - E_b = +E_c > 0
\]

If a label bit was “1”, the corresponding components after the cutoff point in region \(A\) is eliminated to set \(E_b\) to zero,

\[
D = E_c - E_b = -E_b < 0
\]

To extract a label bit from the embedded lc-region, the cutoff point \(c\) of embedding process had to be found back.

Firstly, DEW scheme calculated the energies \(E_c(C_c)\) and \(E_b(C_c)\) to search for cutoff point \(C_c\) in sub-region \(A\) and

\[
\begin{array}{|c|c|}
\hline
\text{SSSS} & \text{AC coefficients} \\
\hline
1 & 1 \\
2 & -3, 2, 2, 3 \\
3 & -7, ..., 4, ..., 7 \\
4 & -15, ..., 8, ..., 15 \\
5 & -31, ..., 16, ..., 16 \\
6 & -63, ..., 32, ..., 32 \\
7 & -127, ..., 64, 64, ..., 127 \\
8 & -255, ..., 128, 128, ..., 255 \\
9 & -511, ..., 256, 256, ..., 511 \\
10 & -1023, ..., 512, 512, ..., 1023 \\
\hline
\end{array}
\]
cutoff point $C_b$ in sub-region $B$:
\[
C_a = \min(C_a), \quad \text{where} \quad C_a = \{ C_i \mid E_i(C_i) < T \} \quad (6)
\]
\[
C_b = \min(C_b), \quad \text{where} \quad C_b = \{ C_i \mid E_i(C_i) < T \} \quad (7)
\]
where $T$ was the detection threshold, which influenced the determination of the cutoff point $c$. It must be smaller than the enforced energy difference $D$ but larger than zero.

Then, DEW regarded the cut off point of embedding as
\[
c = \min(C_a, C_b) \quad (8)
\]
Finally, the extracting procedure was:
- if $C_a < C_b$, the detected watermark bit $W=1$.
- if $(C_a = C_b)$ and $(E_i(C_a) < E_i(C_b))$, $W=1$.
- Otherwise, the detected watermark bit $W=0$.

III. SIGNAL LABELING BY DIFFERENTIAL ENERGY MODULATION

In this section, we firstly suggests that the differential energy $DE$ act as a dither host signal to embed a label bit.

A. Embedded Positions and Differential Energy Definition

When selecting positions for label embedding in DCT domain, we know that DC component of a DCT block represents the average brightness of that block. A change of DC component usually brings a variation to the block luminance and results in block effects. Artifacts will probably propagate to other blocks and the subsequent frames. Although watermarking on DC component of some images may generate much larger perceptual capacity and cause no block effect, its adaptation to all images is still to be confirmed. So we abandon the DC option and select 10 AC coefficients (5 of vertical components and 5 of horizontal ones) neighboring DC component to embed a watermark bit, and label them as $VF_j$ and $HF_j$ ($j=1,2,3,4,5$) respectively.

Fig.2 depicts the selected $VF_j$ and $HF_j$ positions on a DCT block. Since $VF_j$ and $HF_j$ relate to vertical and horizontal changes in low frequency, they have stronger robustness and probability of non-zero values [6].

For calculation convenience and real time applications, we define a simple vertical factor ($VF$) and a horizontal factor ($HF$) to substitute energies of $VF_j$ and $HF_j$ regions.
\[
VF = \frac{1}{5} \sum_{j=1}^{5} |V F_j| \quad (9)
\]
\[
HF = \frac{1}{5} \sum_{j=1}^{5} |H F_j| \quad (10)
\]

In fact, $VF$ and $HF$ are the summation of coefficient magnitudes which represent the distributions of low frequency energy in vertical and horizontal directions. To some extent, the difference $DF$ between $VF$ and $HF$ could approximate the differential energy between vertical texture and horizontal texture of a block.
\[
DF = VF - HF \quad (11)
\]

B. Dither Modulation to Differential Energy

Since $DF$ value is defined on definite positions, the estimation of optimal cutoff point and the evaluation of number of blocks used for one bit carrying could be avoided. By applying DM procedure, we dither $DF$ values to produce a $DF^*$ signal indicating watermark bits. The dithered $DF^*$ sequence is created as follows:

(1) If the embedded bit $w=1$, $DF^*$ is acquired by
\[
DF^* = \begin{cases} 
DF/\Delta & \text{if } DF/\Delta \text{ is even} \\
(DF/\Delta + 1)\times \Delta, & \text{if } DF/\Delta \text{ is odd}
\end{cases} \quad (12)
\]
(2) If the bit $w=0$, $DF^*$ is acquired by
\[
DF^* = \begin{cases} 
(DF/\Delta + 1)\times \Delta, & \text{if } DF/\Delta \text{ is even} \\
DF/\Delta \times \Delta, & \text{if } DF/\Delta \text{ is odd}
\end{cases} \quad (13)
\]

After generating the dither version of $DF$, we modify components of $VF_j$ or $HF_j$ to satisfy the dithered value $DF^*$. The modification involves slight changes of horizontal and vertical coefficients. The dither procedure is:

(1) Calculate the difference between $DF$ and $DF^*$
\[
diff = DF^* - DF \quad (14)
\]
(2) Split $diff$ into two values $val1$ and $val2$
\[
val1 = \text{int}(diff / 2) + 1 \quad (15)
\]
\[
val2 = \text{int}(diff - val1) \quad (16)
\]
(3) Modify coefficients in the embedded sub-region.
Let $HF_{fmax}$ and $VF_{fmax}$ be the biggest value and the second biggest one of $|HF_j|$, $VF_{fmax}$ and $VF_{fmax}$ are
similar defined on \(|\,VF\,|\), they are modified according to the sign of \(\text{diff}\) and the differential energy \(DF\).

- \(\text{diff} > 0\), decrease \(HF\) when \(|HF_{*s}\max| \geq \text{val2}\) or \(|HF_{*f}\max| \geq |\text{diff}|\), otherwise increase \(VF\). Then, \(DF\) will be increased to the nearest quantized value \(DF'\).

- \(\text{diff} < 0\), decrease \(VF\) when \(|VF_{*s}\max| \geq \text{val2}\) or \(|VF_{*f}\max| \geq |\text{diff}|\), otherwise increase \(HF\). \(DF\) will finally be decreased to the nearest quantized value \(DF'\).

- \(\text{diff} = 0\), \(DF = DF'\), no component needs to be changed.

Fig. 3 is the dithering sketch when \(\text{diff} > 0\). We could find out that the suggested DEM algorithm changes DEs in a small range and spreads the change error over coefficients conserved in their positions. Thus it can protect the video well, satisfy rate keeping requirement and be robust to attacks.

C. Watermarking Scheme

(1) Label Bit Embedding

When a watermark sequence \(w_j\) \((j = 0, 1, 2, ..., L-1)\) is embedded in I-frames by DEM, each \(w_j\) is hidden in a quantized DCT block. Suppose an I frame contains \(K\) blocks \((K \geq L)\), at least \(L\) blocks could be selected to label the watermark. The embedding procedure is described as follows.

Step 1: Compute blocked DEs for watermark bit dithering.

Diff. energy \(DF\) of a block is calculated by (11).

\[DF = \text{Project a DE to a consistent dither value.}\]

\(DF\) is modulated to its dither value \(DF'\) by DEM according to the current watermark bit. The difference \(\text{diff}\) between \(DF\) and \(DF'\) is calculated for coefficient adjustment.

Step 3: Coefficient adjustment in a lc-region.

Select horizontal and vertical coefficients in a \(lc\)-region, modify their values to make the energy difference identical to the dither value \(DF'\). \(DF'\) could reflect the watermark bit when decoding.

(2) Watermark Extraction

An embedded watermark signal could be decoded without original video content. The extraction steps are:

Step 1: Pick out a watermarked block, and compute its DF.

Step 2: Calculate the rounded \(\lfloor DF/\Delta \rfloor\).

If it is even, the decoded bit takes value 1, otherwise, the embedded bit is 0.

Step 3: Compare similarities between the original watermark and the extracted label by Normalized cross-correlation function (NC).

\[NC = \frac{(\sum_{j=0}^{l-1} w_j w_j') - (\sum_{j=0}^{l-1} w_j^2)}{\sqrt{(\sum_{j=0}^{l-1} w_j^2 - (\sum_{j=0}^{l-1} w_j^2)^2)}} \]  

where, \(w_j\) and \(w_j'\) are the original watermark and the extracted signal, if a \(NC\) value exceeds a pre-defined threshold, the extracted signal is viewed as an acceptable watermark.

IV. EXPERIMENTAL RESULTS

We use sequences “Forman”, “Carphone” and “Claire” to test the performance of DEM, and compare DEM algorithm with DEW in rate control ability, visual quality and embedding capacity. The test sequences contain 100 frames of size 176x144, with 9 I frames, each frame contains \(K=440\) luminance blocks. For DEM, each I frame can be embedded 440 watermark bits. The step size \(\Delta\) will take values of 3, 4 and 5 respectively. Since in DEW, a watermark bit is represented by a \(lc\)-region which is \(n = 8\times8\) blocks, here we take its typical value \(n=16\). For DEM and DEW, we both embed 25 bits in every I frame. The enforced energy difference in embedding is set to 500, the detection threshold \(T\) is set to 50, and the minimum cutoff point is 6 (i.e. \(C_{\text{min}}=6\)). The experimental results are given as follows.

- Rate Control. Watermarking may change the total bit rate. TABLE II shows a comparison of the maximum and minimum bit rate of the original and the watermarked stream. For the sequences “Forman”, “Carphone” and “Claire”, the biggest variation of the maximum bit rate between the original sequence and the DEM sequence is 600bps \((\Delta = 4)\), 1000bps \((\Delta = 5)\) and 600bps \((\Delta = 4)\) respectively, while the corresponding value of DEW sequence is 1800bps, 1000bps and 1800 bps. And the biggest variation of minimum bit rate between the original sequence and our DEM sequence is 400bps \((\Delta = 3, 4, 5)\), 200bps \((\Delta = 3)\) and 200bps \((\Delta = 5)\) respectively, while the corresponding value of DEW sequence is 800bps, 200bps and 1000 bps. It is obvious that DEM is better than DEW scheme in satisfying the requirement of rate maintenance.

<table>
<thead>
<tr>
<th>Video sequence</th>
<th>Video sequence</th>
<th>Bit rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forman</td>
<td>Original</td>
<td>42200</td>
</tr>
<tr>
<td></td>
<td>DEM (\Delta=3)</td>
<td>421800</td>
</tr>
<tr>
<td></td>
<td>DEM (\Delta=4)</td>
<td>421800</td>
</tr>
<tr>
<td></td>
<td>DEM (\Delta=5)</td>
<td>421400</td>
</tr>
<tr>
<td>Carphone</td>
<td>Original</td>
<td>42000</td>
</tr>
<tr>
<td></td>
<td>DEM (\Delta=3)</td>
<td>421600</td>
</tr>
<tr>
<td></td>
<td>DEM (\Delta=4)</td>
<td>421800</td>
</tr>
<tr>
<td></td>
<td>DEM (\Delta=5)</td>
<td>421400</td>
</tr>
<tr>
<td>Claire</td>
<td>Original</td>
<td>436400</td>
</tr>
<tr>
<td></td>
<td>DEM (\Delta=3)</td>
<td>436800</td>
</tr>
<tr>
<td></td>
<td>DEM (\Delta=4)</td>
<td>437000</td>
</tr>
<tr>
<td></td>
<td>DEM (\Delta=5)</td>
<td>436800</td>
</tr>
</tbody>
</table>

- Transparency. Visual quality is evaluated by the average degradation of PSNR. Fig. 4 gives PSNR curves of the original and the watermarked “Forman” stream respectively. In our DEM scheme, the maximum degradation of PSNR in I frame is not bigger than 1dB when \(\Delta = 3, 4, 5\). In fact \(\Delta = 4\) or 5 is a large step size for a QIM quantized signal. In addition, the suggested DEM algorithm spreads the amount of change over
selected elements of a lc-region. Thus $\Delta = 3$ is enough for embedding. Furthermore, PSNR curves show that quality of P/B frames is maintained, and the average degradation of the sequence is 0.046dB when $\Delta = 3$, 0.071 dB when $\Delta = 4$, and 0.0690 when $\Delta = 5$. In contrast, the average degradation of DEW with the same capacity is more than 1.5dB in I frame and the average degradation of the sequence is 0.204dB. Thus DEM is superior to DEW in preserving visual quality when the same label bits are embedded.

• Robustness. Table III gives the NC values of “Foreman” sequence after attacks to test the DEM performance of frame dropping attack resistance. It demonstrates that DEM is robust to frame dropping. All the NC values of DEM and DEW are 1 when dropping 5~15 P/B frames. For dropping 2~6 I frames, the NC values of DEM are 1 when $\Delta = 3$–5, while NC values of DEW decrease from 1.000 to 0.375.

• Capacity. The capacity of a block-based watermarking method is determined by the number of blocks used to embed one label bit. For DEW, one bit is usually embedded in $n \times 8$ blocks. Ref.[2] suggested that $n$ satisfy $n = 16 \times k^2$, where $k=1,2,3, \ldots$, its typical value is 16–64. As $n$ becomes larger, PSNR values increase, and the watermark more robust, but the capacity decreases. For DEM method, one watermark bit is embedded in one block with better visual quality than DEW, its DM mechanism could project the original value to a correct quantized position by just changing it slightly. Thus DEM has 16 times greater capacity than DEW. Moreover, the redundant capacity can provide more robustness to attacks by repeated embedding. To emphasize the test results, we also gives some experimental images in Fig.5.

V. CONCLUSION

This paper proposes a new video watermarking scheme in compress domain, based on local energy dithering. It employs DM to modulate differential energy of a block. Simulation results show that our scheme satisfies the requirement of rate maintenance better than DEW. In our scheme, a watermark bit can be embedded in a block in I frame by only slightly changing some non-zero AC coefficients. Thus, DEM achieves greater high perceptual and capacity than DEW.

ACKNOWLEDGEMENT

The authors acknowledge the kindly supports of China NNSF (Grant No: 60472063) and GNDSF (04020074).

REFERENCES


<table>
<thead>
<tr>
<th>TABLE III</th>
<th>NC VALUES AFTER ATTACKS</th>
</tr>
</thead>
<tbody>
<tr>
<td>attacks</td>
<td>DEW</td>
</tr>
<tr>
<td>without attacks</td>
<td>1.000</td>
</tr>
<tr>
<td>number of dropped P/B frames</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1.000</td>
</tr>
<tr>
<td>10</td>
<td>1.000</td>
</tr>
<tr>
<td>15</td>
<td>1.000</td>
</tr>
<tr>
<td>number of dropped I frames</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1.000</td>
</tr>
<tr>
<td>3</td>
<td>0.938</td>
</tr>
<tr>
<td>4</td>
<td>0.875</td>
</tr>
<tr>
<td>5</td>
<td>0.500</td>
</tr>
<tr>
<td>6</td>
<td>0.375</td>
</tr>
</tbody>
</table>