Highly Imperceptible Video Watermarking with the Watson’s DCT-based Visual Model

Han-Min Tsai, Long-Wen Chang
Department of Computer Science,
National Tsing Hua University, Hsing-Chu, Taiwan, R.O.C
E-mail: hmtsai@cs.nthu.edu.tw; lchang@cs.nthu.edu.tw

Abstract

Digital watermarking plays an important role for copyright protection, content authentication, and annotation of digital multimedia. To embed watermark quickly, we embed the watermark in the compressed domain instead of the frequency domain. For high imperceptibility, a new video watermarking based on Human Visual System (HVS) is proposed. It embeds the watermark in the block that is highly tolerant of noise.

The proposed algorithm directly substitutes the Variable Length Code (VLC) without inverse quantization and forward quantization. It can greatly reduce computational complexity. In the experiment, we apply MPEG-1 video sequences as our test sequence and the results demonstrate the high imperceptibility of watermark.

1. Introduction

The digital watermarking technique has become a research interest due to concerns about illegal piracy of copyrighted content [1]. Podilchuk and Zeng proposed perceptually based watermarking schemes [3, 4]. They derive the maximum strength watermark without introducing perceptual difference. Their schemes are based on optimizing DCT quantization matrices for each image. The optimized DCT quantization matrices hinge on image independent frequency sensitivity, image dependent luminance sensitivity and contrast masking by utilizing a visual model in the context of image compression.

Langelaar et al. [5, 6] proposed two labeling methods, which embed the label information directly into an MPEG compressed video bitstream. The first method [5] embeds the label by changing the specific variable length codes. The second method [6] discards some of the high frequency DCT coefficients in the bitstream with a probability mass function. In this paper, we will utilize the low-complexity property of the first method but we do not take the second method into account due to its expensive computation.

2. The Watson’s DCT-based visual model

Watson’s DCT-based visual model [8] estimates the perceptibility of changes in individual terms of image’s block DCT, and then pools those estimates into a single estimate of perceptual distance. This model first defines a frequency sensitivity table. Each table entry \( t[i, j] \), is approximately the smallest magnitude of the corresponding DCT coefficient in a block that is discernible in the absence of any masking noise (i.e., the amount of change in the coefficient that produces one Just Noticeable Difference (JND). The fact that the magnitude in the left-top block is small implies the eye is more sensitive to the frequency.

When this model derives luminance masked threshold, Watson mentions that brighter regions of an image are able to absorb large changes without becoming noticeable. Therefore, the proposed technique embeds a watermark signal to these brighter regions without introducing noticeable difference.

Based on the above discussion, we apply the Watson’s contrast masked threshold to derive contrast masked thresholds for each block in a Y picture of the I frame in a color video sequence. We compute the summation values of contrast masked thresholds for each block. We suggest that those blocks with high summation value are capable of carrying large additional energy and resisting attacks.

3. The proposed embedding algorithm

We apply Watson’s DCT-based visual model to estimates each blocks’ luminance and contrast masked
threshold in the Y picture of the I frame in a color video sequence. The capacity of a block is then defined as summation of contrast masked threshold in a block. The blocks with high capacity will be chosen as the host blocks for embedding the watermark, and VLC codes in host blocks can be substituted in the watermark embedding process. The following steps explain overall embedding scheme illustrated in Figure 2 in a color video sequence and Figure 1 illustrates the block-embedding scheme.

**Step 1. Divide the Y picture into a sequence of 8x8 blocks and derive the luminance masked threshold and contrast masked threshold for each DCT frequency in the block.**

The Watson’s DCT-based visual model [8] defines the luminance masked threshold for each DCT frequency in a 8x8 block, as given by

$$t_L[i,j,k] = (i,j)(C_o[0,0,k] / C_{o,0})^{a_T}$$  \hspace{1cm} (1)

where $a_T$ is a constant with a suggest value of 0.649, $C_o[0,0,k]$ is the DC coefficient of the $k^{th}$ block in a picture, $t_{L[i,j]}$ is the DCT frequency sensitivity as shown in Table 1, $C_{o,0}$ is the average among the DC coefficients in a picture, and $N$ is the block numbers of a Y picture. The dimension of the Y picture in our experiment is 352 x 240, so $N$ is (352/8) x (240/8) =1320. From equation (1), the luminance masked threshold will be large if the DC coefficient is large. This explains brighter blocks of an image will be able to absorb larger changes without becoming noticeable.

**Table 1. The DCT frequency sensitivity table**

<table>
<thead>
<tr>
<th>1.40</th>
<th>1.01</th>
<th>1.16</th>
<th>1.66</th>
<th>2.40</th>
<th>3.43</th>
<th>4.79</th>
<th>6.56</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.01</td>
<td>1.45</td>
<td>1.32</td>
<td>1.52</td>
<td>2.00</td>
<td>2.71</td>
<td>3.67</td>
<td>4.93</td>
</tr>
<tr>
<td>1.16</td>
<td>1.32</td>
<td>2.24</td>
<td>2.59</td>
<td>2.98</td>
<td>3.64</td>
<td>4.60</td>
<td>5.68</td>
</tr>
<tr>
<td>1.66</td>
<td>1.52</td>
<td>2.56</td>
<td>3.77</td>
<td>4.25</td>
<td>5.30</td>
<td>6.28</td>
<td>7.60</td>
</tr>
<tr>
<td>2.40</td>
<td>2.00</td>
<td>2.98</td>
<td>4.55</td>
<td>6.15</td>
<td>7.46</td>
<td>8.71</td>
<td>10.17</td>
</tr>
<tr>
<td>3.43</td>
<td>2.71</td>
<td>3.64</td>
<td>5.30</td>
<td>7.46</td>
<td>9.62</td>
<td>11.58</td>
<td>13.51</td>
</tr>
<tr>
<td>4.79</td>
<td>3.67</td>
<td>4.60</td>
<td>6.28</td>
<td>8.71</td>
<td>11.58</td>
<td>14.50</td>
<td>17.29</td>
</tr>
<tr>
<td>6.58</td>
<td>4.93</td>
<td>6.68</td>
<td>7.60</td>
<td>10.17</td>
<td>13.51</td>
<td>17.29</td>
<td>21.15</td>
</tr>
</tbody>
</table>

The Watson’s DCT-based visual model [8] then defines the contrast masked threshold for each DCT frequency in the block, as given by

$$s[i,j,k] = \max[i,j,k] C_o[i,j,k]^{0.7} t_L[i,j,k]^{0.3}$$  \hspace{1cm} (2)

where $t_L[i,j,k]$ is the luminance masked threshold for each term of the block DCT, and $C_o[i,j,k]$ is the each DCT coefficient in the block $k$. The formula (2) clearly decides that the contrast threshold value depends on not only the energy present in that frequency but also the luminance masked threshold for that frequency. The Watson’s model also indicates that the final thresholds, $s[i,j,k]$, estimate the amounts by which individual terms of the block DCT may be changed before resulting in one JND.

**Figure 1. Embedding the watermark in the host block**

**Step 2. Calculate the capacity for each block and the mean capacity in a Y picture.**

We define the capacity of a block is the summation of contrast masked threshold in the block. For each block, we calculate the capacity of the block $k$, given by

$$S_k = \sum_{i=1}^{8} \sum_{j=1}^{8} s[i,j,k] \quad (3)$$

where $s[i,j,k]$ is the contrast masked threshold of $i,j^{th}$ position of the block $k$.

After we calculate each block capacity in a Y picture, we can have the max capacity among these blocks in a Y picture, denoted by $S_{MAX}$. Meanwhile, we derived the mean capacity of these blocks in a Y picture, as given by

$$S_m = (1 / N) \sum_{k=1}^{N} S_k \quad (4)$$

where $N$ is the block numbers of a Y picture, and $S_k$ is the capacity for the block $k$. 

![Diagram](diagram.jpg)
Step 3. Find the host block.

If \( S_k > (1 / 2)(S_m + S_{\text{MAX}}) \) (5)

The block \( k \) is chosen as the host block, where \( S_m \) is the mean capacity in a Y picture, and \( S_{\text{MAX}} \) is the max capacity among these blocks in a Y picture.

Step 4. Embed the Watermark in the host block.

Applying Huffman decoding to those VLC code in the host block, we restore the original run-level pairs. To embed a watermark bit \( W[i] \), we subsequently increase the \( L \) value with strength \( \alpha \), as given by

\[
L^* = L + \alpha \cdot W[i], \quad 0 \leq i \leq 4095
\]  

(6)

where \( W[i] \) is the pixel of bi-level image in line-scan order (i.e. \( W[i] \) is 1 if the pixel is white, \( W[i] \) is 0 if pixels is black).

According to the VLC code table for MPEG-1, we obtain the new VLC code for modified run-level pair. Subsequently, we directly substitute the original VLC codes for new VLC ones. Therefore, the watermark bits are embedded on those nonzero AC coefficients in the zigzag scan order, as illustrated in Figure 1.

Additionally, the equation (2) has many unnecessary computations due to that many AC coefficients \( C_{[i,j,k]} \) are zero after VLC decoding of the host block. In other words, we can set \( s[i,j,k] = t_{[i,j,k]} \) in most cases. To do this, we first calculate the contrast masked thresholds of those points with nonzero AC coefficients. For the other points with zero AC coefficients, we can reduce the computations by directly assigning \( t_{[i,j,k]} \) to \( s[i,j,k] \).

4. The extracting algorithm

In the extraction process, we locate the embedded positions in the original video stream and the watermark-embedded video stream and extract the watermark \( W^* \) according to formula (7).

\[
W^*[i] = \begin{cases} 1 & \text{if } L^*_i - L_i > 0 \\ 0 & \text{otherwise} \end{cases}
\]  

(7)

where \( L^*_i \) is the AC value in the \( i^\text{th} \) embedded position of embedded video stream, \( L_i \) is the AC value in the \( i^\text{th} \) embedded position of original video stream and \( W^*_i \) is the \( i^\text{th} \) extracted watermark bit. We verify the extracted watermark by measuring similarity between the extracted watermark and the original one, as given by

\[
\text{Sim}(W, W^*) = \frac{\sum_{i=0}^{4095} W[i] \oplus W^*[i]}{\sum_{i=0}^{4095} W[i] \oplus W[i]}
\]  

(8)

5. Experiment results

We embed the watermark in the host blocks (Figure 3.a) and the blocks without considering visual model (Figure 3.c) with the same number bits ’1’ and \( \alpha = 1 \). As anticipated, embedding watermark in the host blocks has better fidelity than embedding blocks without considering visual model. This experiment also indicates that the host block is capable of absorbing large change without introducing perceptual difference. We find out those embedded positions lie on high frequency blocks or brighter blocks (Figure 3.d) and human visual system has less sensitivity to the noise in the host blocks.

Figure 4 shows that if we embed 1024 bits in Y picture of football video sequence, we can have PSNR= 43.30 with \( \alpha = 1 \) and PSNR=32.01 with \( \alpha = 5 \). Figure 5 shows that if we embed \( \alpha = 5 \) in a Y picture of football video sequence, we can have PSNR=40.60 with 128 bits and PSNR= 32.01 with 1024 bits. It gives us two choices to achieve high imperceptibility (1) by using proper embedding bits under fixed \( \alpha \) or (2) by using proper \( \alpha \) under fixed embedding bits.

![Figure 2. The overall embedding scheme](image-url)
6. Conclusion

In this paper, we proposed a novel watermarking scheme for a compressed video sequence such as MPEG-1. Comparing with those frequency-domain-based watermarking schemes that must perform VLC decoding, inverse quantization, watermark embedding, forward quantization, and VLC encoding, we only do VLC decoding and VLC code substitution. We also successfully applied Watson’s DCT-based visual model to have better imperceptibility for video watermarking.

![Figure 3](image)

(a) The embedded Y picture with considering visual model. (b) The original Y picture. (c) The embedded Y picture without considering visual model. (d) The embedded blocks with considering visual model. (e) The embedded blocks without considering visual model.

![Figure 4](image)

Figure 4. Various $\alpha$ values vs. PSNRs under the fixed embedding bits 1024.

![Figure 5](image)

Figure 5. Various Bits vs. PSNRs under fixed $\alpha = 5$.

7. References

[5] Gerrit C. Langelaar, Reginald L. Lagendijk, Jan Biemond “Real-time labeling of MPEG-2 compressed video” Faculty of information technology and system(ITS), information and communication theory group, delft university of technology
[8] Ingemar J. Cox Matthew L. Miller Jeffery A. Bloom; DIGITAL WATERMARKING, p215-p218