VIDEO WATERMARKING USING TEMPORAL SENSITIVITIES OF HUMAN VISUAL SYSTEM

Alper Kız and A. Aydın Alatan

Department of Electrical and Electronics Engineering.
M.E.T.U., Balgat, 06531, Ankara, Turkey.
Phone: +90 312 210 23 51; Fax: +90 312 210 12 61
e_mail: koz@metu.edu.tr, alatan@ece.metu.edu.tr

ABSTRACT

A video watermarking method is presented, based on the temporal sensitivity of Human Visual System (HVS). The method exploits the temporal contrast thresholds of HVS to determine the spatio-temporal locations, where the watermark should be embedded, and the maximum strength of watermark, which still gives imperceptible distortion after watermark insertion. The robustness results indicate that the proposed scheme survives video distortions, such as additive Gaussian noise, ITU H.263+ coding at medium bit rates, frame dropping and frame averaging without any perceptual artifacts.

Keywords: Human Visual System, Temporal Contrast Thresholds, ITU H263+, watermarking.

1. INTRODUCTION

Digital watermarking has emerged in the recent years as a solution to the problem of making secure electronic data, which may provide the solid proof of ownership. Researchers working on this area focused on the requirements to provide useful and effective watermarks [3-6]. The requirements for an effective watermark are imperceptibility, robustness to any signal processing operations, and capacity that refers to the ability of detecting the watermark among different watermarks with a low probability of error. The imperceptibility criterion is directly and the other two are indirectly related to the human visual system (HVS)[4]. Hence, the researchers working on digital watermarking usually utilize basics of HVS such as contrast masking [1], spatial masking [2] and foveation [3] for image watermarking.

However, in video watermarking, the adaptation of HVS is more difficult compared to the image data. There are two main reasons in video watermarking that makes the problem different. The first one is the necessity of a guaranteed robustness against new type of attacks, such as frame dropping, frame averaging, and collusion, which have no counterparts in image watermarking. The second one is due to the temporal imperceptibility of the watermark, which is relatively more difficult problem, due to 3-D characteristics of video. In order to provide an imperceptible watermark, the watermarking procedure should also take the variations in the temporal direction into account.

In order to solve the mentioned problems, Wolfgang, et al [4] proposed an HVS-based video watermarking method that achieves a trade off between independent watermarking each frame and utilization of the same watermark for each frame. The method embeds a watermark into each intra (I) frame in an MPEG sequence using the DCT based perceptual image watermarking method [4] and then, applies a simple linear interpolation to the watermarks for every frame between two consecutive I-frames. In another study [6], a novel shot-based video watermarking method is proposed. The video is separated into shots for obtaining temporal stationary signals and temporal wavelet transform is computed for each shot. A different watermark is embedded to each wavelet coefficient frame by exploiting the contrast masking and spatial masking characteristics of HVS [6].

While the spatial properties of HVS are widely used in watermarking research, its temporal counterparts have not been utilized in the literature. In this study, we propose an alternative method that exploits the temporal contrast thresholds [7,8] of HVS. In the experiments determining the thresholds [7], the visual target of specific spatial frequency, \((u,v)\), is modulated by a sinusoidal time-domain function of a specific temporal frequency, \(w\). The temporal contrast threshold, \(T(u,v,w)\), is the contrast of the modulating sinusoid function when the temporal fluctuations in the visual target become visible [7]. Therefore, by the definition, the modifications, which are smaller than the temporal contrast thresholds in the temporal direction of the visual target, should be invisible. In other words, temporal contrast thresholds determine the maximum level of the watermark that will be embedded into the video towards temporal direction.

2. PROPOSED WATERMARKING METHOD

The overall structure of the watermarking procedure is given in Fig.1 (a). The first step is to separate the video into its shots. For each shot, the intensities are then converted into its contrast values. Contrast [7] is
Fig. 1 (a) Overall structure of the watermark insertion process, (b) Watermark detection

\[
C'(bx, by, u, v, w) = \left\{ \begin{array}{l} 
C(bx, by, u, v, w) + W(bx, by, u, v, w)T(u, v, w) \quad \text{if } |C(bx, by, u, v, w)| \leq T(u, v, w) \\
|C(bx, by, u, v, w)| \quad \text{otherwise}
\end{array} \right.
\]
correlation is first found for each discrete temporal frequency \( w \) and then, the mean in \( w \) direction is taken:

\[
W(bx, by, u, v, w) = \frac{1}{w} \sum_{w=1}^{w} W(bx, by, u, v, w)
\]

\[
p(w) = \frac{1}{w} \sum_{w=1}^{w} \text{inner product of } v_{1, 2}
\]

(5)

Finally, the mean is compared to a threshold for detection.

![Figure 2 Frames from Coast, Carphone and Pingpong video.](image)

(a) (b) (c) (d) (e) (f)

**Figure 2** Frames from Coast, Carphone and Pingpong video. (a), (c) and (e) are original frames, (b), (d) and (f) are watermarked frames. PSNR values are 39.9, 39.6 and 41.4 dB, respectively.

### 4. SIMULATION RESULTS

**Coastguard**, **Carphone** and **Pingpong** sequences (176x144) are utilized during simulations. Only the first 60 frames of these sequences are used and the watermark is embedded only into Y-component.

Typical samples from original and watermarked frames for each video are illustrated in Fig. 2. The original and watermarked frames are visually indistinguishable. However, the visual equivalence of the watermarked and original frames does not require the visual equivalence of the watermarked and original sequences. As noted, the differences between the watermarked and original sequences might become visible due to the temporal characteristics of the video. Due to this situation, the watermarked and original video are presented to a number of subjects and tested whether they perceive the difference between two video. According to these informal tests, the videos are assumed as visually equal.

According to the motion content of the video, the number of the coefficients to be watermarked may differ. The graph of the number of the watermarked coefficients with respect to discrete temporal frequency is illustrated in Fig 3. The number of watermarked coefficients is decreasing due to the increase in \( T(u, v, w) \) and decrease in the magnitude of the temporal discrete Fourier transform of the sequence while the temporal frequency is increasing. In Fig. 4, the magnitude of the difference between the temporal discrete Fourier transform of the original and watermarked video sequences are illustrated for 4 different discrete frequency, i.e., magnitude of \( (C^w(bx, by, u, v, w) - C(bx, by, u, v, w)) \) is illustrated for 4 different \( w \). As \( w \) increases, the watermarked coefficients are the ones that correspond to the high motion regions of the video. For \( w = 0 \), (DC case), most of the low spatial frequency elements of the 8x8 blocks are watermarked.

![Figure 3 The number of watermarked coefficients vs. discrete temporal frequency for Carphone qcf sequence. Total number of watermarked coefficients is 45563 for Carphone sequence, 70270 for Coastguard sequence and 72404 for Pingpong sequence.](image)

During simulations for testing robustness, same operations are applied to both original and watermarked video. Watermark embedding and detection processes are repeated 30 times with different watermarks for each case. The minimum value of these correlation results, when watermark is present in the video, and the maximum value of the correlation results, while watermark is not present are determined. The higher separation between that minimum and maximum values shows the robustness of the system.

### 4.1 Robustness to Additive Gaussian Noise

In order to model video compression techniques that are based on the temporal sensitivity of HVS (e.g. such as 3-D transform coding), the watermarked video is corrupted with additive Gaussian noise of zero mean and 0.1 variance that is added to the video in the temporal frequency domain after multiplied with temporal contrast thresholds.
In Table 1, the correlation results for each video sequence with and without watermark are tabulated. It is important to note that the minimum correlation values with watermark are much larger than the maximum correlation values without watermark. The mean of the inner product results (see (5)) after 30 runs is drawn as a function of discrete temporal frequency in Fig. 5(a) for each video sequence with and without watermark. It is clearly seen that at each frequency, the difference between the correlations for watermarked and unwatermarked case is quite high.

4.2 Robustness Against ITU H263+ Coding

One of the possible operations on video is a lossy coding stage, applied for the purpose of storage and transmission of digital video at low bit rates. The robustness of the watermarking method against ITU H263+ coding is tested for different bit rates. In the testing process, the bit rate is decreased until 240 kbps. It is observed that the watermark may survive until this bit rate. The correlation results are presented in Table 2. In Fig. 5(b), the inner product results for each different temporal frequency are also shown. While the inner product for the DC term (w = 0) is quite high, the results for the AC terms are becoming lower. This is due to compression, which distorts mostly AC terms. Below 230 kbps, increasing compression rate makes watermark undetectable.

4.3 Robustness Against Frame Dropping and Averaging

Some other distortions, which are based on temporal characteristics of the digital video, are temporal cropping, frame dropping and frame interpolation. An attacker can maintain the visual quality of the digital video by dropping some frames from the video and/or by replacing them by frame interpolation. For the frame dropping case, one of each consecutive two frames is dropped, which can be accepted as one of the worst case while still maintaining the visual quality. For the frame interpolation case, one of each consecutive two frames is dropped and replaced by the average of two neighboring frames. Each of these attacks mainly distorts the high frequency components of the video. Therefore, only the low frequency components (first 15 components) are taken into account while computing the correlation in the detection part. The correlation results for frame dropping and frame averaging are illustrated in Tables 3 and 4, respectively. In Fig. 6(a) and (b), the inner product results for each different temporal frequency are also shown, respectively, for the case of frame dropping and frame averaging. As it is observed in Fig. 6(b), inner product results are decreasing steadily while the frequency increases since the frame averaging distorts mainly the high frequency components.
Figure 5 Mean of the inner product results as a function of the discrete temporal frequency after (a) Gaussian noise and (b) ITU H.263+ coding at a bit rate of 230 kbps, for Coast sequence (‘x’s and ‘o’s show the correlation results for the watermarked and original video, respectively).

5. CONCLUSIONS

A video watermarking approach is proposed based on the temporal sensitivity of HVS. The method embeds the watermark in the temporal Fourier domain by exploiting the temporal contrast thresholds in [8]. The robustness results show that the watermarking scheme can survive the typical video attacks, such as additive Gaussian noise, ITU H263+ coding, frame dropping and frame averaging.

One interesting point for the results is obtaining better robustness of the DC term of video compared to AC terms, especially in the test for ITU H263+ coding. While the correlation for the AC terms of the video cannot be detected after the bit rate of 240kbps, the correlation result for the DC term stands up to lower bit-rates, even 50kbps. One may conclude that ITU H.263+ coding distorts mostly the temporal AC components of video as well as spatial counterparts.

The proposed method embeds the watermark only into Y component of video. Extending the scheme into chromatic components will improve the robustness, while slightly loosing from imperceptibility due to low sensitivity of chromatic components.

Figure 6 Mean of inner product results vs temporal frequency after (a) frame dropping, and (b) frame averaging for Coast.

One other possible extension of the method can be realized by the use of temporal masking phenomenon of HVS [7]. In such a scheme, the temporal contrast threshold for a specific temporal frequency increases due to the masking of a temporal variation at a different temporal frequency. This phenomenon of HVS can be interpreted as the contrast masking [1,9] in temporal direction. One may expect the robustness of such a watermarking scheme will be better compared to the proposed method due to the increase in temporal contrast thresholds.

6. ACKNOWLEDGEMENTS

This research is supported by The Scientific and Technical Research Council of Turkey (TÜBİTAK) under project EEEAG 101E007.

7. REFERENCES


<table>
<thead>
<tr>
<th>Video</th>
<th>PSNR (dB)</th>
<th>With watermark</th>
<th>Without watermark</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Max</td>
<td>Mean</td>
</tr>
<tr>
<td>Coast</td>
<td>26.9</td>
<td>0.9721</td>
<td>0.9670</td>
</tr>
<tr>
<td>Carphone</td>
<td>27.8</td>
<td>0.9822</td>
<td>0.9810</td>
</tr>
<tr>
<td>Pingpong</td>
<td>26.6</td>
<td>0.9633</td>
<td>0.9604</td>
</tr>
</tbody>
</table>

Table 1 Correlation results for Coast, Carphone and Pingpong sequences after Gaussian noise.

<table>
<thead>
<tr>
<th>Video</th>
<th>Bit rate (kbps)</th>
<th>PSNR (dB)</th>
<th>With watermark</th>
<th>Without watermark</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Max</td>
<td>Mean</td>
<td>Min</td>
</tr>
<tr>
<td>Coast</td>
<td>230</td>
<td>29.4</td>
<td>0.1796</td>
<td>0.1416</td>
</tr>
<tr>
<td>Carphone</td>
<td>246</td>
<td>34.5</td>
<td>0.2238</td>
<td>0.2093</td>
</tr>
<tr>
<td>Pingpong</td>
<td>236</td>
<td>33.9</td>
<td>0.2122</td>
<td>0.2042</td>
</tr>
</tbody>
</table>

Table 2 Correlation results for Coast, Carphone and Pingpong sequences after ITU H263 + Coding.

<table>
<thead>
<tr>
<th>Video</th>
<th>With watermark</th>
<th>Without watermark</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Max</td>
<td>Mean</td>
</tr>
<tr>
<td>Coast</td>
<td>0.3983</td>
<td>0.3925</td>
</tr>
<tr>
<td>Carphone</td>
<td>0.2410</td>
<td>0.2287</td>
</tr>
<tr>
<td>Pingpong</td>
<td>0.1570</td>
<td>0.1487</td>
</tr>
</tbody>
</table>

Table 3 Correlation results for Coast, Carphone and Pingpong sequences after frame dropping.

<table>
<thead>
<tr>
<th>Video</th>
<th>With watermark</th>
<th>Without watermark</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Max</td>
<td>Mean</td>
</tr>
<tr>
<td>Coast</td>
<td>0.4295</td>
<td>0.4215</td>
</tr>
<tr>
<td>Carphone</td>
<td>0.2737</td>
<td>0.2624</td>
</tr>
<tr>
<td>Pingpong</td>
<td>0.1693</td>
<td>0.1578</td>
</tr>
</tbody>
</table>

Table 4 Correlation results for Coast, Carphone and Pingpong sequences after frame averaging.