A NEW ALGORITHM FOR COLLUSION RESISTANT VIDEO WATERMARKING

Vinod P[*] *and P. K. Bora* [†]

Department of Electronics and Communication Engineering
Indian Institute of Technology, Guwahati, INDIA 781 039

ABSTRACT

This paper presents a new algorithm for collusion resistant *multi-bit* video watermarking. Video frames from the same scene of the video sequence are decomposed into wavelet coefficient frames by temporal wavelet transform and the watermark is embedded into the low-pass temporal frames. Embedding is done in the DCT domain of the low-pass temporal wavelet frames by applying the perceptual masking properties of the DCT coefficients. The watermark decoder is a blind statistical decoder. In the decoder, the underlying probability distribution of the mid-frequency DCT coefficients of the temporal low-pass wavelet frames is approximated by *Generalized gaussian distribution*. The robustness of the proposed algorithm against different video degradations and distortions is evaluated and presented.

1. INTRODUCTION

One of the challenging issues in video watermarking is its robustness against inter-frame collusion attack. Collusion attack exploits the redundancy in the host data or in the watermark to estimate the redundant component. Most of the video watermarking algorithms proposed so far consider the video as a sequence of still images and apply existing still image watermarking techniques to each frame. Such watermarking schemes are not robust to collusion attack. The pioneering work in collusion resistant video watermarking was done by Swanson *et al.* [1]. In their proposed *multiresolution scene-based* video watermarking technique, the perceptual masking properties of the Human Visual System (HVS)are exploited to embed a highly robust watermark. The video sequence to be watermarked is segmented into different scenes and for each scene, the temporal wavelet transform of the frames in the scene is taken. The watermark is added to the low-pass and high-pass frames of the temporal wavelet transform. Thus the watermark has static and dynamic components and is claimed to have a high degree of robustness to collusion attack [2].

The contrast masking model for the HVS used in the watermark embedding process in [1] was originally proposed for the 8×8 block DCT coefficients of still images. Since the contrast mask is a non-linear function of the DCT coefficients, the contrast mask estimated from the temporal wavelet frame is not *visually optimal*. The watermark detector in [1] is a non-oblivious detector which requires both the unwatermarked video and the original watermark. The use of two transforms, two masking schemes and non-oblivious detector makes the watermarking scheme fairly complex. In this paper, we propose a new video watermarking scheme which eliminates the aforementioned drawbacks. The watermark is embedded in the mid-frequency DCT coefficients of the low pass temporal wavelet frames. The watermark decoder is a blind statistical decoder which does not require the original video sequence. The major contributions of this work includes: extending the temporal wavelet based scheme [1] to blind watermarking scheme, More accurate exploitation of the visual masking property, Better statistical modelling of the host data and reduced computational complexity.

[*] e-mail: vinod@iitg.ernet.in
[†] e-mail: prabin@iitg.ernet.in

2. BACKGROUND

2.1. Scene based multiresolution video watermarking

The multiresolution scene-based watermarking proposed by Swanson *et. al.* [1] exploits the spatial masking, frequency masking and the temporal properties to embed an invisible and robust watermark. The watermark consists of temporally *static* and *dynamic* components and can be detected even from a single frame of the video scene without knowing the temporal placement of the frame in the video sequence. The video sequence to be watermarked are first segmented into different scenes. To each frames in a scene, the temporal wavelet transform is applied and the resulting temporal low-pass and high-pass frames are 8×8 block DCT transformed to get the DCT frames. The contrast mask is then estimated from these DCT coefficients. A key dependant pseudorandom sequence which represents the owner of the video is generated as a DCT-domain signal. This psuedorandom sequence is perceptually shaped with the frequency mask. IDCT is applied to the resulting sequence, then multiplied with the spatial mask to get the watermark and is added to the temporal wavelet frames. The resulting wavelet frames with the embedded watermark are then converted back to the temporal domain using the inverse temporal wavelet transform. The watermark detection process is non-oblivious, in which both the original video and the watermark are required. Depending on the availability and the knowledge of the test frame placement relative to the original video, two types of detection hypothesis have been proposed.

2.2. DCT-domain still Image watermarking

In [3], a structure for DCT-domain multi-bit watermarking and theoretical basis for the performance analysis of the detector have been proposed. The perceptual masking model is only luminance based, disregarding the contrast masking in the Watson's model [4] due to the oblivious nature of the watermark detector. The luminance mask $T(i,\ j)$ for an 8×8 DCT block is given by,

$$T'(i,j) = T(i,j)\left(\frac{X_{0,0}}{\overline{X}_{0,0}}\right)^{a_T} \quad i,\ j \in 0,\ 1, \cdots,\ 7. \quad (1)$$

0-7803-8554-3/04/$20.00 ©2004 IEEE.

where, $a_T = 0.649$, is a constant, $X_{0,0}$ is the DC-DCT coefficient of the block, $\overline{X}_{0,0}$ is the average luminance of the display (1024 for an 8-bit image). The final perceptual mask $\alpha[k_1, k_2]$, which gives the maximum is permissible alteration to the 8×8 block DCT coefficient $X[k_1, k_2]$ of an image is given by

$$\begin{aligned} \alpha[k_1, k_2] &= 4 \cdot \left(1 + \left(\sqrt{2} - 1\right) \delta(l_1)\right) \\ &\quad \cdot \left(1 + \left(\sqrt{2} - 1\right) \delta(l_2)\right) \cdot \gamma \cdot T'(l_1, l_2) \end{aligned} \tag{2}$$

where $l_1 = k_1 \bmod 8$, $l_2 = k_2 \bmod 8$, $\delta(\cdot)$ is the kronecker function and $\gamma < 1$ is a scaling factor. As a compromise between the robustness of the watermark and the visual quality of the watermarked image, the watermark is added only to the mid-frequency DCT coefficients of each 8×8 block. The watermark carries a multi-bit hidden message.

The watermark verification process consists of two steps. First presence of watermark in a given test image is detected and then if it contains a watermark, the hidden message is decoded. Both the detection and the decoding steps are oblivious and based on the *Maximum Likelihood* principle where each 8×8 DCT coefficient is considered as a sample of an independent identically distributed (iid) generalized Gaussian stochastic process with probability density function

$$f_x(x) = Ae^{-|\beta x|^c}. \tag{3}$$

where,

$$\beta = \frac{1}{\sigma} \left(\frac{\Gamma\left(\frac{3}{c}\right)}{\Gamma\left(\frac{1}{c}\right)} \right)^{\frac{1}{2}}, \qquad A = \frac{\beta c}{2\Gamma\left(\frac{1}{c}\right)}$$

$\Gamma(\cdot)$ is the gama function, σ is the standard deviation and c is a constant.

3. PROPOSED ALGORITHM

Similar to the method proposed by Swanson *et al.* [1], the proposed method achieves collusion resistance by embedding watermark in the temporal wavelet frames of each scene. Due to large volume of data, digital video is always stored and manipulated in the compressed form. All the video compression techniques try to reduce the temporal redundancy in the video frames and hence mainly affect the high pass temporal components of the video. So in the present work, the watermark is embedded only in the low pass temporal wavelet frames of the video. Each lowpass temporal wavelet frame is watermarked with still image watermarking technique. This present work uses one of the well known DCT-domain still image watermarking technique proposed by Hernandez *et. al.* [3] for watermarking the temporal wavelet frames. That is, the watermark is embedded into the mid-frequency DCT coefficients of the temporal wavelet frames.

Since both the DCT and wavelet transform are linear transforms, the order in which they are applied during the watermark embedding process can be interchanged. If the temporal wavelet transform(TWT) is first applied to the video frames, the TWT for all the pixels have to be calculated followed by the DCT computation of the temporal low-pass frames. Conversely, the DCT can be first applied to all the video frames and then only the mid-frequency DCT coefficients which are to be modified during the watermarking process are subjected to TWT. The second approach *ie.*, first the DCT and then TWT, is used in the proposed algorithm due to the perceptual masking model which will be explained in the following section. In addition, this approach will reduce the number of TWT computations and hence the overall computational complexity of the watermarking scheme because TWT is computationally intensive than DCT.

3.1. Perceptual masking

Due to the oblivious nature of the proposed watermarking scheme, it uses the luminance based perceptual masking model explained in subsection 2.2. As mentioned earlier, the watermark is embedded to the DCT coefficients of the TWT frames and so the perceptual mask has to be calculated for the DCT of TWT farmes. Since the perceptual masking model is defined for the DCT coefficients of the gray scale images, and is a nonlinear function of the DC-DCT coefficients of each 8×8 block, it cannot be applied to the temporal wavelet frames as done in [1]. But we can take the advantage of the linearity of DCT and temporal wavelet transforms to get the perceptual mask. This can be done by first calculating the luminance mask for each DCT frame of the video sequence, and then finding the temporal wavelet transform of the luminance mask for each frame. Unfortunately this will double the number of temporal wavelet transform computations required. This additional computational burden can be eliminated by using the fact that the luminance mask for an 8×8 DCT block depends only on the DC value of the block. The luminance mask given in Eqn.[2] can be rewritten as:

$$\alpha[k_1, k_2] = \tau(k_1, k_2) \left(X_{0,0}\right)^{a_T} \tag{4}$$

where,

$$\begin{aligned} \tau(k_1, k_2) &= 4 \cdot \left(1 + \left(\sqrt{2} - 1\right) \delta(l_1)\right) \\ &\quad \cdot \left(1 + \left(\sqrt{2} - 1\right) \delta(l_2)\right) \cdot \gamma \cdot \frac{T(l_1, l_2)}{\left(\overline{X}_{0,0}\right)^{a_T}} \end{aligned} \tag{5}$$

Since $\tau(k_1, k_2)$ is image independent and same for all the video frames, it can be taken as a constant while finding the temporal wavelet transform of the luminance mask and hence we need to find only the temporal wavelet transform of $(X_{0,0})^{a_T}$. The resulting temporal wavelet coefficients are multiplied with $\tau(k_1, k_2)$ to get the perceptual mask.

3.2. Watermark encoding

Consider a scene of L frames each of size $N_1 \times N_2$, denoted by $x_i[\mathbf{n}] = x_i[n_1, n_2]$, $0 \le n_1 < N_1$, $0 \le n_2 < N_2$, $i = 1, 2, \cdots, L$. Let $S = \{\mathbf{s}\}$, $D = \{\mathbf{d}\}$ be respectively the sets of points in the discrete grid corresponding to the mid-frequency DCT coefficients and the DC coefficients of all the blocks in a frame. The N-bit hidden message M carried by the watermark be, $\mathbf{b} = (b_1, b_2, \cdots, b_N)$, where $b_i \in \{-1, 1\}$. Diiferent steps involved in the watermark encodinng process is given in Figure 1. The following are the steps:

1. To each frames, apply the 8×8 block DCT to get the DCT frames
2. Apply the TWT to the mid-frequency DCT coefficients $X_i[\mathbf{s}]$ to get the wavelet frames $Y_i[\mathbf{s}]$. Let L_p be the number of low-pass frames.
3. Find the perceptual mask $M_i[\mathbf{s}]$, $i = 1, \cdots, L_p$ for each low-pass temporal wavelet frame as described in the preceding section

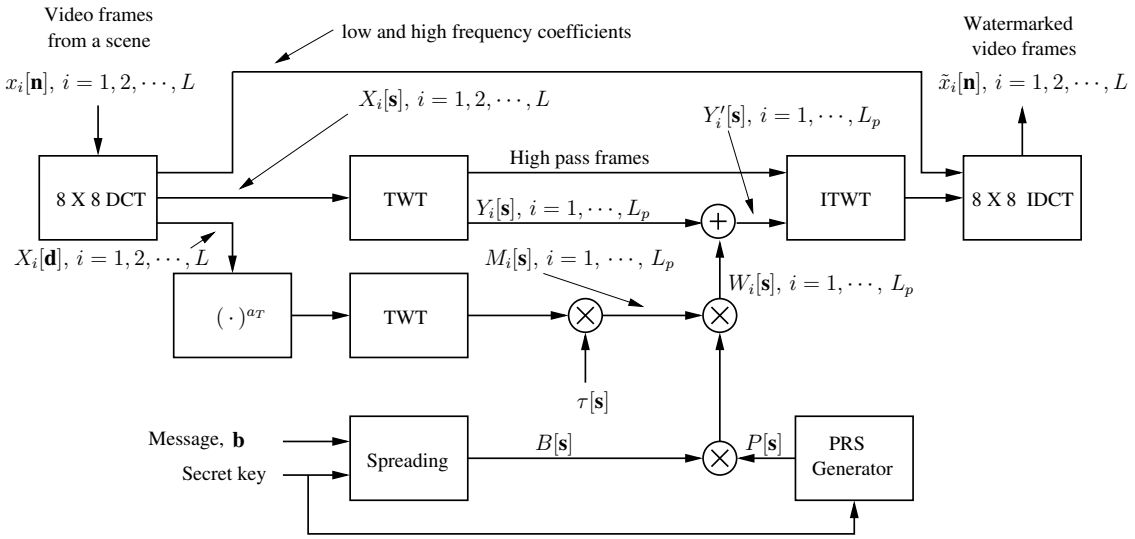

Fig. 1. Watermark Encoder

4. Divide the set S into N non-overlapping patitions $S_i,\ i = 1,\ 2, \cdots,\ N$. Generate a 2-D sequence $B[\mathbf{s}]$ as given by

$$B[\mathbf{s}] = b_i, \quad \forall \mathbf{s} \in S_i \quad i = 1,\ 2\ \cdots,\ N$$

5. A 2-D sequence $P[\mathbf{s}]$,whose elements are from the set $\{-1,\ 1\}$, is generated using a secret key.

6. The watermark for each temporal low-pass wavelet frame is calculated as

$$W_i[\mathbf{s}] = M_i[\mathbf{s}] \cdot B[\mathbf{s}] \cdot P[\mathbf{s}], \quad i = 1,\ 2\ \cdots,\ L_p$$

7. The watermark is added to the low-pass temporal wavelets frame as

$$Y_i'[\mathbf{s}] = Y_i[\mathbf{s}] + W_i[\mathbf{s}], \quad i = 1,\ 2\ \cdots,\ L_p$$

8. The watermarked low-pass wavelet frames together with the high pass frames are subjected to inverse temporal wavelet transform(ITWT) and then the resulting sequence together with the low and high frequency DCT coefficients of each DCT frame are subjected to 8×8 block IDCT to get the watermarked video frames $\tilde{x}_i[\mathbf{n}],\ i = 1,\ 2,\ \cdots,\ L$.

3.3. Watermark Decoder

In this work, we are considering only the hidden message decoding process, assuming that the test video contains a watermark with a given key. The watermark decoder is an oblivious statistical decoder based on the *Maximum likelyhood* principle. In blind watermark decoding, since the host data is not available, it has to be statistically modelled. Most of the watermarking techniques proposed so far, use the gaussian assumption for the host data and in that case the optimal decoder is the simple correlation based decoder. But better statistical modelling like *Generalized Gaussian Distribution* for mid-frequency 8×8 DCT coefficients[3] has considerably increased the performance of the oblivious watermark detectors as compared to the conventional correlation detector. Such a statistical model for the host data is not yet proposed in the case of video watermarking. In this work, the mid-frequency DCT coefficients for each temporal low pass frame are approximated by the Generalized Gaussian Distribution given in Eqn.3, with a different standard deviation σ for each low pass frame.

Using the above statistical model and the procedure given in [3], it can be shown that the coefficients, $r_{i,\,j}\ i = 1,\ 2,\ \cdots, N$ given by,

$$r_{i,\,j} = \sum_{\mathbf{s} \in S_i} \left|Y_j'[\mathbf{s}] + M_j[\mathbf{s}] \cdot P[\mathbf{s}]\right|^c - \left|Y_j'[\mathbf{s}] - M_j[\mathbf{s}] \cdot P[\mathbf{s}]\right|^c \quad (6)$$

are sufficient statistics for the optimum decoding of the message bit b_i from the j$^{\text{th}}$ low pass temporal wavelet frame. The decoded value of the bit b_i embedded in the j$^{\text{th}}$ low pass frame is given by,

$$\hat{b}_i = sgn(r_{i,\,j}),\ i \in \{1,\ 2, \cdots,\ N\},\ j \in \{1,\ 2, \cdots,\ L_p\} \quad (7)$$

The message embedded in each temporal low pass frame is decoded by using the Eqn. 7. The final decoded message is obtained by majority voting of each decoded bit in all the low pass temporal wavelet frames.

3.4. Experimental results

In order to evaluate the performance of the algorithm, a number of experiments were carried out on 32 frames of four standard gray-level test video sequences "Football", "Tempete", "Mobile" and "Table tennis". Two level temporal wavelet transform of the DCT frames carried out using filters corresponding to "Daubechies 3" wavelets. The watermarks were embedded in 22 mid-frequency DCT coefficients as in [3] of the resulting 8 low pass temporal frames using the procedure given in Section 3.2. For getting the results in a reasonable simulation time, the perceptual masks $M_i[\mathbf{s}]$ were multiplied with a scaling factor less than one.The average PSNR of the watermarked sequences are 48.86 , 48.53 , 48.69 and $49.26\ dB$ respectively for the "Football", "Tempete","Mobile" and "Table tennis" sequences.

In the decoder, the mid-frequency DCT coefficients of each temporal low pass wavelet frames were modelled as Generalized

	JPEG quality factor	*BER*
JPEG compression	20	0.1870
	30	0.1110
	40	0.0560
	50	0.0310
	compression ratio	*BER*
MPEG compression	100:1	0.0790
	125:1	0.1150
	150:1	0.1940
	175:1	0.2110
	standard deviation of noise	*BER*
Additive gaussian noise	25.50	0.0210
	36.06	0.0340
	44.16	0.0550
	51.00	0.0740
	57.02	0.0810
	62.46	0.0920

Table 1. *BER performance*

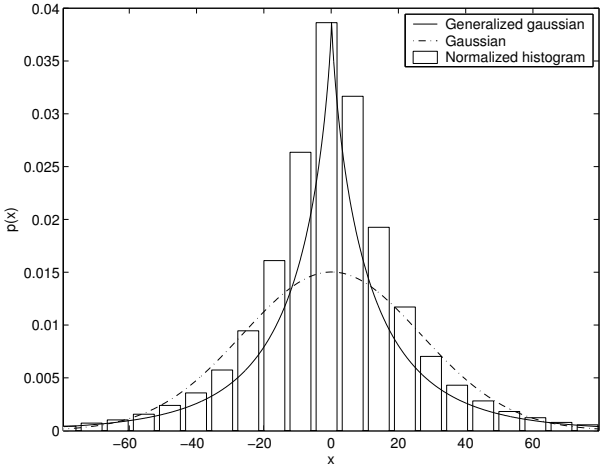

Fig. 2. Normalized Histogram of mid-frequency DCT coefficients of low pass temporal wavelet frame and statistical modelling

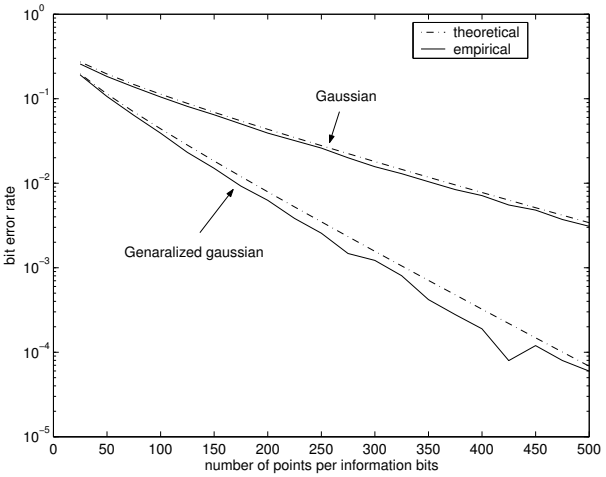

Fig. 3. BER as a function of number of points in S_i

Gaussian distribution and following [3], the value of c was taken as 0.8. Figure. 2 shows the statistical modelling of the mid-frequency DCT coefficients of one temporal low pass frame of the "mobile" sequence. Figure. 3 shows the the theoretical and empirical average *bit error rate* (BER) performance of the proposed decoder and the conventional correlation based decoder as a function of the number points in the set S_i. The theoretical BER curves were obtained by assuming the proposed statistical model and following the procedure given in [3]. From Figures 2 and 3 it is clear that the proposed decoder better models the host data and hence shows the increased BER performance. In order to evaluate the robustness of the proposed algorithm to various watermarking attacks like MPEG compression, JPEG compression of each watermarked frames, and additive gaussian noise, a number of experiments were carried out. In all the experiments, the video sequences were watermarked with watermarks carrying 100 bits of information. The experiments were repeated for 10 different keys and the average BER performance for "table tennis" sequence is given in Table 1. The results show the robustness of the proposed algorithm to different watermark attacks.

4. CONCLUSIONS

A new algorithm for multi-bit watermarking of uncompressed video is presented. The proposed algorithm is robust against inter-frame collusion attack because the watermark is embedded in the temporal wavelet frames. The algorithm uses a simplified and *visually optimal* perceptual masking model without additional computational burden. The watermark decoder assumes a Generalized Gaussian model for the host data and is oblivious. Experimental results show the improved performance of the proposed decoder as compared to the conventional correlation based decoder and the robustness of the watermark to different watermark attacks.

5. REFERENCES

[1] M. D. Swanson, B. Zhu and A. H. Tewfik, "Multiresolution Scene-based Video Watermarking Using Perceptual Models," *IEEE Journal of Selected Areas in Communications*, vol. 16, pp. 540–550, May 1998.

[2] K. Su, D. Kundur, and D. Hatzinakos, "A Novel Approach to Collusion-Resistant Video Watermarking," in *Security and Watermarking of Multimedia Contents IV. E. J. Delp and P. W. Wong, eds.,*, January 2002, vol. 4675.

[3] Juan R Hernandez, Martin Amado, and Fernando Perez-Gonzalez, "DCT-Domain Watermarking Techniques for Still Images: Detector Performance Analysis and a New Structure," *IEEE Transactions on Image Processing*, vol. 9, pp. 55–68, January 2000.

[4] A. B. Watson, "Visual optimization of DCT quantization matrices for individual images," in *Proc.AIAA Computing in Aerospace 9*, 1993, pp. 286–291.