Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.12785/amis/072L17

Video Watermarking Using Spatio-Temporal Masking For ST-DM

Hongbo BI[1,2], *Yubo ZHANG*[2] *and Xueming LI*[1]

[1]Beijing Key Laboratory of Network System and Network Culture, Beijing 100876, China
[2]School of Electrical Information Engineering, NorthEast Petroleum University, Daqing 163318, China

Received: 3 Oct. 2012, Revised: 15 Feb. 2013, Accepted: 18 Feb. 2013
Published online: 1 Jun. 2013

Abstract: Traditional Spread Transform Dither Modulation (ST-DM) is based on the fixed quantization step size, which does not exploit the features of the video contents. Spatio-temporal masking function represents the visual threshold under which human eyes are usually not sensitive. In this paper, we adopts not only the spatial contrast sensitivity function (CSF), the luminance masking function, the contrast masking function but also the spatio-temporal masking function to account for the model of human eyes to the video. we propose using spatio-temporal perceptual model to adaptively select the quantization step size based on the calculated masking function. Experimental results show that the proposed scheme guarantees high peak signal to noise ratio (PSNR). Furthermore, compared with the scheme with fixed step size, the proposed scheme is more robust against temporal frame averaging (TFA) attack.

Keywords: Video Watermarking, Spatio-Temporal Masking, ST-DM, Temporal Collusion Attack

1 Introduction

With the rapid development of computer and network technology, more and more videos propagate over Internet in the form of digital data, which causes the issue of copyright protection [1]. Digital video watermarking is one of the most efficient solutions to prevent illegal copy, distribution of video, which embeds the copyright information into the host video without degrading the video greatly.

Watermarked video may encounter various attacks during transmission, including intentional or unintentional ones [2]. Among of them, temporal collusion is one of the most difficult attacks, which may destroy the watermarking system.

Generally, we can classify two types of temporal collusion, inter-video collusion and inter-frame collusion. In the first case, each different version of the same video contains different watermark. The colluders can average all the frames so as to weaken the strength of each different watermark, without introducing any great visual distortion. Temporal Frame Attack (TFA) is common in this category. In the second type of temporal collusion, each frame of the same video contains same watermark. The colluders can estimate the watermark by averaging a sufficiently large number of frames. Then the colluder can exploit the knowledge of the watermark to remove it from the host video simply by subtraction. Watermark Estimation and Remodulation attack (WER) is among of this category [3].

Robustness against collusion attacks has been studied in spread spectrum algorithms [4][5][6]. It has been shown that robustness against TFA requires that the same watermarking signal is embedded in consecutive frames.

Quantization Index Modulation (QIM) attracts more eyes due to the good rate-distortion-robustness performance [7]. Similar analysis of robustness against TFA with spread spectrum algorithms has ever been done for the case of Spread Transform Dither Modulation (ST-DM) watermarking [8]. On the other hand, video watermarking commonly needs to meet the visual imperceptibility. A variety of algorithms have been proposed to improve the imperceptibility by different Human Visual System (HVS) [9]. However, no similar work has been done in ST-DM. We move the first step into this direction by introducing the spatio-temporal masking function, which simultaneously improves the visual imperceptibility and robustness against the TFA as it will be proved.

* Corresponding author e-mail: bhbdq@126.com

This paper is organized as follows. In section 2, inter-video collusion attack (IVCA), especially the TFA is presented, we analyze the temporal averaging operation of the video. In section 3, we formula the ST-DM algorithm. In section 4, Spatio-temporal Just Noticeable Distortion (JND) Profile is proposed, we give the spatio-temporal masking function. In section 5, we propose the ST-DM scheme with adaptive quantization step size, which preserves the video quality and provides the robustness against TFA. In section 6, some experimental results are mentioned. Finally, conclusions are given in section 7.

2 Inter-Video Collusion Attacks (IVCA)

In the inter-video collusion, the colluder may make use of several watermarked frames. For example, the colluder may average several consecutive frames within a video shot to produce a video sequence from which it is no more possible to extract the watermark. This kind of attack, commonly named Temporal Frame Averaging (TFA) is clearly efficient if consecutive frames contain same watermark.

TFA can be considered as the temporal averaging between consecutive video frames where the colluder replaces each watermarked frame $f_w(t)$ with an averaged frame computed as follows

$$\overline{f}_w(t) = \frac{1}{W}\sum_{k=1}^{W} f_w(t+k) \tag{1}$$

where W represents the number of frames used for the averaging.

The TFA attack is viewed as a low-pass filtering in purpose to remove the watermarking by preserving the visual quality of the attacked video. It is well known that the averaging operation provides smooth value by discarding the outlier values.

3 Spread Transform Dither Modulation (ST-DM)

Digital watermarking can be modeled as communication with side information. Quantization index modulation (QIM) provides a efficient method. QIM also provides a computational efficient watermarking algorithm implementing the blind extraction. However, the standard QIM algorithm employs a fixed quantization step size which may lead to poor fidelity. A number of improved algorithms have been proposed in recent years including adaptive QIM and adaptive RDM both using the Watson distance.

The ST-DM scheme was first introduced by Chen and Wornel. The fundamental idea of this scheme is to spread the watermark into several host signals. The diagraph of the algorithm is shown in Figure 1.

More specifically, the ST-DM first projects the host signal v onto the vector v^{st} along the direction u and then the result is quantized before being added to the host signal which is orthogonal to u.

$$v^{w\_st} = v + (Q(v^T u, \Delta, b, \xi) - v^T u)u, b \in \{0,1\} \tag{2}$$

Theoretically, the vector u can be arbitrarily chosen only if it satisfies the condition $< u,u >= 1$, where $< \cdot,\cdot >$ represents the inner product. Recenly in [10], Braci etal. proposed that the better robustness against TFA can be achieved by exploiting the orthogonal vector such as Hadamard matrix.

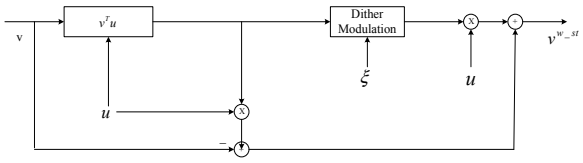

Fig. 1: ST-DM

The hard decision rule is expressed as follows:

$$\hat{b} = \arg\min_{w\in\{0,1\}} \left| v^{w\_st}u - Q(v^{w\_st}u, \Delta, w, \xi) \right| \tag{3}$$

And the corresponding soft decision is

$$\hat{b} = \arg\min_{w\in\{0,1\}} \sum_{i=1}^{M} \left| v_i^{rw\_st}u - Q(v_i^{rw\_st}u, \Delta, w, \xi) \right| \tag{4}$$

4 Spatio-Temporal Just Noticeable Distortion (JND) Profile

The visual perceptual model of the human visual system (HVS) should account for luminance masking, contrast masking, sensitivity and temporal sensitivity etc.. Spatio-temporal Just noticeable difference (JND) is considered as the minimum threshold that is generally perceptible visually, and is sometimes used to measure the distance between two videos.

The Just Noticeable Distortion (JND) in the DCT domain can be typically expressed as the product of a spatial masking function and a temporal masking function [11].

$$T_{st}(k,m,n,i,j) = T_s(k,m,n,i,j) \cdot T_t(k,m,n,i,j) \tag{5}$$

$$T_s(k,m,n,i,j) = T_{sb}(k,m,n,i,j) \cdot T_{sm}(k,m,n,i,j) \tag{6}$$

$$T_{sm}(k,m,n,i,j) = T_l(k,m,n,i,j) \cdot T_c(k,m,n,i,j) \tag{7}$$

Where T_{st}, T_s and T_t are the spatio-temproal JND threshold, the spatial masking function and the temporal

masking function, respectively. Since each frame is partitioned into blocks, k represents the frame index, m, n represent the block on the m^{th} row n^{th} column of the frame, and i , j represent the pixel on the i^{th} row j^{th} column within the block.

T_{sb} is the spatial base threshold which is generated by the spatial contrast sensitivity function (CSF) which considers the spatial summation effect factor and the oblique effect factor first introduced in Ahumada's JND model.

The base threshold for a DCT subband is expressed as

$$T_{sb}(k,m,n,i,j) = s \cdot T_b(m,n,i,j) \tag{8}$$

$$T_b(k,m,n,i,j) = \frac{1}{\phi_i \phi_j} \cdot \frac{1}{r+(1-r)\cdot\cos^2\varphi_{ij}} \cdot \exp(c \cdot w_{ij})/(a+b \cdot w_{ij}) \tag{9}$$

Where parameter s accounts for the spatial summation effect and takes the value of 0.25. ϕ_i and ϕ_j are DCT normalization factors

$$\phi_m = \begin{cases} \sqrt{\frac{1}{N}} , & m = 0 \\ \sqrt{\frac{2}{N}} , & m > 0 \end{cases} \tag{10}$$

The term $\frac{1}{r+(1-r)\cdot\cos^2\varphi_{ij}}$ accounts for the oblique effect, where r is set to 0.6 and φ_{ij} stands for the directional angle of the corresponding DCT component

$$\varphi_{ij} = \arcsin\left(\frac{2\omega_{i,0}\omega_{0,j}}{\omega_{ij}^2}\right) \tag{11}$$

$$\omega_{ij} = \frac{1}{2N}\sqrt{(i/\theta_x)^2 + (j/\theta_y)^2} \tag{12}$$

$$\theta_x = \theta_y = 2 \cdot \arctan\left(\frac{1}{2 \times R_{vd} \times Pic_h}\right) \tag{13}$$

Where N is the dimension of the DCT block (8 in our experiments), θ_x and θ_y are the horizontal and vertical visual angles of a pixel. Here, R_{vd} stands for the ratio of viewing distance to picture height. Pic_h is the number of pixels in picture height.

In [11], a psychophysical perceptual experi-ment was performed to test the JND threshold to determine the parameters a, b and c which are set to 1.33, 0.11 and 0.18, respectively.

Furthermore, $T_{sm}(k,m,n,i,j)$ is the product of the luminance masking function and the contrast masking function in [11].

According to the Weber's law, an experimental formula of the luminance masking function is adopted. Contrast masking function is computed based on the classification of regions by Canny operator, where the blocks can be classified into three categories, namely plane, edge and texture, respectively.

Finally, the temporal masking function is expressed as:

$$T_t(k,m,n,i,j) = \begin{cases} 1 & , f_s < 5cpd, f_t < 10Hz \\ 1.07^{(f_t-10)} & , f_s < 5cpd, f_t \geq 10Hz \\ 1.07^{f_t} & , f_s \geq 5cpd \end{cases} \tag{14}$$

Where

$$f_t = f_{sx}v_x + f_{sy}v_y \tag{15}$$

$$f_{sx} = \frac{i}{2N\theta_x} , \ f_{sy} = \frac{j}{2N\theta_y} \tag{16}$$

$$v_\zeta = v_{I\zeta} - v_{E\zeta} , \ (\zeta = x,y) \tag{17}$$

$$v_{E\zeta} = \min[g_{spem} \times v_{I\zeta} + v_{MIN}, v_{MAX}] \tag{18}$$

Where g_{spem} is the gain of the smooth pursuit eye movements with the empirical value of 0.98. v_{MIN} is the minimum eye velocity due to the drift movement and the classical value is 0.15 deg/s. v_{MAX} is the maximum velocity of the eyes corresponding to the saccadic eye movement and the value is normally 80 deg/s.

$$v_{I\zeta} = f_{fr} \times MV_\zeta \times \theta_\zeta \quad (\zeta = x,y) \tag{19}$$

Where f_{fr} is the frame rate of video sequence. MV_ζ is the motion vector of each block, which is obtained by the three-step-search motion estimation in this paper. θ_ζ is the visual angle of a pixel.

Finally, the masking function accounts for the spatio-temporal characteristics of the video [12].

5 Proposed Watermarking Scheme

Generally, video watermarking application should meet the two most important and mutually conflicting requirements — visual imperceptibility and robustness. The spatio-temporal masking function provides a potential solution for video watermarking to improve the visual imperceptibility. By locally changing the quantization step size according to the video contents, we are able to provide significantly improved visual quality. For example, in regions of high texture, a larger step size can be used, while in regions of low texture, a small step size is chosen. As we shall demonstrate, this adap-tivity can also simultaneously improve robustness. Furthermore, the quantization scheme assures the blind extraction, which is crucial to the video watermarking.

The overall structure of the proposed watermarking scheme is depicted in Figure 2.

The proposed method mainly consists of two parts. In the first part, the video sequence is partitioned into scenes and transformed into the 3D-DCT domain. In the second part, the spatio-temporal masking function is adopted to determine the embedding quantization step, and the watermark bit is embedded into the 3D-DCT coefficients adaptively.

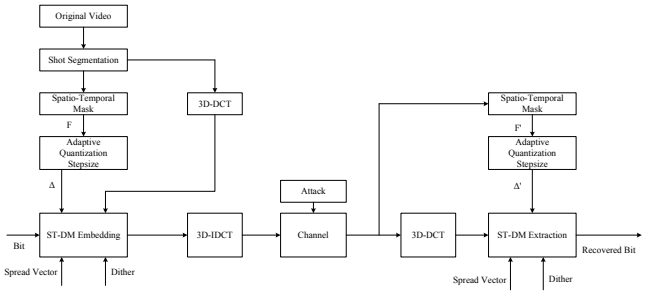

Fig. 2: Block diagraph of the proposed scheme

The detailed embedding steps are described as follows:

Step 1. Video shot segmentation

The video is segmented into consecutive scenes in order to obtain temporally stationary signals. This decomposition yields a more suitable framework for exploiting the spatio-temporal JND profile. Then, each scene is partitioned into the temporal frames segment selected by secret key. For the sake of simplicity, we have segmented the video sequence into scenes of fixed length.

Step 2. 3D-DCT

Each resulting frame is segmented into non-overlapping blocks, where coefficients are transformed into the 3D-DCT domain by means of spatial 2-D DCT transform followed by a DCT transform in the temporal domain.

Step 3. Spatio-temporal masking function

In our scheme, we compute the spatio-temporal masking function in which we adopt luminance masking, contrast masking, sensitivity and temporal sensitivity etc. accounting for the static and dynamic objects in the video.

Step 4. Adaptive quantization step size

The spatio-temporal masking function is multiplied by a global constant to determine the quantization step size for each 3D-DCT coefficient. The global constant is also known to the receiver. Note that, we also proceed the spread transform to the spatio-temporal masking function so as to obtain the final quantization step size.

Step 5. ST-DM Embedding

We perform the ST-DM scheme where the spread direction (projection vector) is random.

Step 6. Reconstruction of watermarked video

For each non-overlapping block, repeat the step 2, 3, 4, 5 until the whole watermark is embedded into the host frame. Then perform the inverse 3D-DCT combined with unchanged transform coefficients, we obtain the watermarked frame.

The extraction of the watermark is very simple and it is a blind procedure. For the simplicity of analysis, we do not consider the case that the colluder also could add noise when performing the TFA. The first 4 steps are same to those in embedding. The watermark extraction is hard decision because of the single bit embedding. It is necessary to point out that due to the possible attack, the watermarked video at the receiver may be different from that before the transmission, as a result, the spatio-temporal masking function and final step size are different from those at the transmitter, which causes some extraction errors. However, this kind of error is tiny due to the shot segmentation as proved in the experiments.

6 Experimental Results

For assessing the performance of the proposed algorithm, a variety of experiments were carried out. The video is Bus_cif with size of 288*352, where the video contains 150 frames. The block size for embedding is 8*8, and the size of the watermark is 36*44 bits.

The 4th frame and the watermarked frame are shown in Figure 3. The peak signal to noise ratio (PSNR) is nearly 35.36 dB, and the watermarked frame maintains good visual fidelity, so it seems difficult to distinguish the host and the watermarked frame. Furthermore, the bit error ratio is 0, which reveals we can extract the watermark accurately.

Fig. 3: Embedding the watermark (a) Original video frame, left (b) Watermarked video frame, right

To evaluate the robustness against TFA of the proposed scheme further, Figure 4 and Figure 5 illustrate results after TFA attacks on the watermarked frame which indicates the proposed scheme is much more robust than the one using the fixed quantization step size.

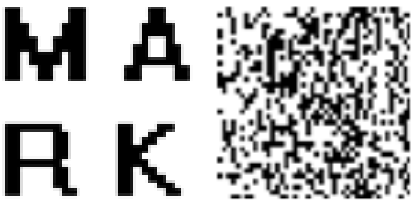

Fig. 4: Extracting the watermark (fixed step size) (a) Original watermark, left (b) Recovered watermark, right

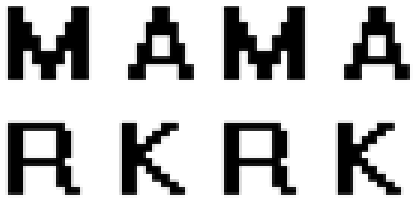

Fig. 5: Extracting the watermark (proposed scheme) (a) Original watermark, left (b) Recovered watermark, right

7 Conclusions and Future work

In this paper, masking function in spatio-temporal domain is employed in the video watermarking based on ST-DM scheme. Spatio-temporal masking function embodies the visual threshold under which the change of video contents can not be perceptible by eyes while introducing tiny degradation of visual quality. Applying spatio-temporal masking function to video watermarking is beneficial to improve the visual imperceptibility. The experimental results show that the proposed scheme improves not only the high perceptual quality, but also is robust against TFA. The key idea of the proposed algorithm is introducing the masking function by spread transform the 3D-DCT coefficients to obtain the adaptive quantization step size. Further work of integrating the characteristics of HVS algorithms and robustness against more attacks into our approach is in progress.

Acknowledgement

This work is supported by the Science & Technology Project of Heilongjiang province under Grant No. 12521056.
The authors are grateful to the anonymous referee for a careful checking of the details and for helpful comments that improved this paper.

References

[1] Hongbo BI, Yubo ZHANG, Xueming LI, Video watermarking robust against spatio-temporal attacks, Journal of Networks, vol. 6, no. 6, pp. 932-936, June 2011.
[2] Pik Wah Chan, Michael R. Lyu, Roland T. Chin. A Novel Scheme for Hybrid Digital Video Watermarking: Approach, Evaluation and Experimentation. IEEE TRANSACTION ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, 2005, vol.15, no.12, pp: 1638-1649.
[3] G. Doërr, J.-L. Dugelay, A guide tour of video watermarking, Signal Process.: Image Commun., vol. 18, pp. 263-282, Apr. 2003.
[4] Karen Su, Deepa Kundur, Dimitrios Hatzinakos, Statistical Invisibility for Collusion-Resistant Digital Video Watermarking, IEEE TRANSACTIONS ON MULTIMEDIA, vol.7, no. 1, pp. 43-51, Feb. 2005.
[5] Karen Su, Deepa Kundur, Dimitrios Hatzinakos, Spatially localized image-dependent watermarking for statistical invisibility and collusion resistance, IEEE TRANSACTIONS ON MULTIMEDIA, vol. 7, no. 1, pp. 52-66, Feb. 2005.
[6] Tomáš KANÓCZ, Tomáš TOKÁR, Dušan LEVICKÝ, Robust frame by frame video watermarking resistant against collusion attacks, RADIOELEKTRONIKA 2009. 19th International Conference, pp. 99-102.
[7] B. Chen and G.W. Wornell, Quantization Index Modulation: A Class of Provably Good Methods for Digital Watermarking and Information Embedding, IEEE Transactions on Information Theory, vol. 47, no. 4, pp. 1423-1443, 2001.
[8] R. Caldelli, A. Piva, M. Barni, A. Carboni, Effectiveness of ST-DM Watermarking Against Intra-video Collusion, Lecture Notes in Computer Science, vol. 3710, pp. 158-170, september 2005.
[9] Qiao Li, Ingemar J. Cox, Using Perceptual Models to Improve Fidelity and Provide Resistance to Valumetric Scaling for Quantization Index Modulation Watermarking, IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, VOL. 2, NO. 2, pp. 127-139, JUNE 2007.
[10] Sofiane Braci, Remy Boyer, Claude Delpha, ANALYSIS OF THE RESISTANCE OF THE SPREAD TRANS-FORM AGAINST TEMPORAL FRAME AVERAGING ATTACK, ICIP2010, pp. 213-216.
[11] Zhenyu Wei, King N. Ngan, Spatio-Temporal Just Noticeable Distortion Profile for Grey Scale Image/Video in DCT Domain, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 19, NO. 3, pp. 337-346, MARCH 2009.
[12] Yubo ZHANG, Hongbo BI, Transparent Video Watermarking Exploiting Spatio-Temporal Masking in 3D-DCT Domain, Journal of Computational and Information Systems, vol. 7, no. 5, pp. 1706- 1713, May 2011.

Hongbo BI received his bachelor degree in 2001 and master degree in 2004 respectively from NorthEast Petroleum University both in communications engineering. Currently he is a vice professor in School of Electrical Information Engineering in NorthEast Petroleum University. Now he is a doctorial candidate at School of Information and Communication Engineering in Beijing University of Posts and Telecommunications. His main research interests focus on digital watermarking, signal processing, digital video processing, multimedia telecommunications.

Yubo ZHANG received her bachelor degree in 2004 and master degree in 2007 respectively from NorthEast Petroleum University both in communications engineering. Currently she is a lecturer in School of Electrical Information Engineering in NorthEast Petroleum University. Her main research interests focus on digital watermarking, signal processing, multimedia telecommunications. She has published research articles in international journals.

Xueming LI received his B.S degree from University of Science and Technology of China in 1992 and Ph.D. degree from Beijing University of Posts and Telecommunications in 1997, all in electronics engineering. From 1997 to 1999, he was a post-doctor researcher in Institute of Information Science of Beijing Jiaotong University. Currently he is a professor at School of Information and Communication in Beijing University of Posts and Telecommunications. In 2002, he once worked as a guest lecturer in Karlsruhe University, Germany. His current research interests include digital image processing, video coding, and multimedia telecommunications.