SECURE AND HVS-ADAPTIVE EXHIBITION SPREAD TRANSFORM DITHER MODULATION WATERMARKING FOR DIGITAL CINEMA

Rony Darazi, Pilar Callau and Benoît Macq

Communications and Remote Sensing laboratory, Université catholique de Louvain
2 Place du Levant, B-1348, Louvain-la-Neuve, Belgium
phone: + (32) 10 47 23 00, fax: + (32) 47 20 89, email: Rony.Darazi@uclouvain.be
web: www.tele.ucl.ac.be

ABSTRACT

In this paper, we propose a secure watermarking scheme based on Spread Transform Dither Modulation (STDM) method for Digital Cinema. The embedding is performed in the JPEG2000 decoding pipeline after the de-quantization and prior to the inverse discrete wavelet transform (IDWT). We exploit the wavelet properties related to the Human Visual System (HVS) in order to have a trade-off between Fidelity and Robustness, while preserving Security. We design a pixel-wise masking vector that modulate the spreading vector in such a way that preserve its security. Our results show that the proposed method is robust against traditional image processing attacks. The proposed scheme can also survive the camecording attack, a pre-processing step is done in the detection for this end.

***Index Terms*—** Exhibition Watermark, Quantization Index Modulation, Perceptual Masking, JPEG2000, Digital Cinema

1. INTRODUCTION

We can distinguish two main classes for images watermarking techniques: the additive class known by spread spectrum (SS) methods introduced by Cox et al. [1], and the substitutive class known by Quantization Index Modulation (QIM) methods introduced by Chen and Wornell [2]. In addition to their high capacity-robustness trade-off, QIM methods are simple and have a small computational cost.

Spread Transform Dither Modulation (STDM) is a variant of QIM and is mainly robust against re-quantization.

A lot of interest has been payed to approaches that use the Human Visual System (HVS) properties to adapt the watermark. These approaches differ in the way they fix the watermarking strength. The maximum amount of change that can be tolerated before the human eye detects a difference or (JND), is determined in those methods by the exploitation of HVS perceptual characteristics.

In [3], Barni et al. presented a SS watermarking method, where the watermark is embedded in the DWT-domain through pixel-wise masking. In this approach, the image is decomposed into sub-levels and the low-frequency coefficients are watermarked by using a controlled quantization process. More recently, Li and Cox presented a STDM watermarking algorithm in the DCT-domain using Watson's perceptual models [4]. In this method, for each DCT coefficient, they define a "Slack" which refers to the quantity by which the DCT coefficient can be altered. In [5], Yu et al. proposed the same algorithm with an improved luminance-masked threshold. Note that QIM methods are largely vulnerable to rounding operation in lossy compression, re-quantization and valumetric scaling.

In Digital Cinema specifications, JPEG2000 a DWT-based coder is adopted. Moreover, the watermark insertion is recommended to perform in the decoding pipeline which is done in realtime.

In this work, a secure and adaptive low-complexity exhibition watermarking method is proposed to fit in the JPEG2000 decoding pipeline, prior to the inverse wavelet transform, and such that it has minimal impact on image quality. Our main contribution is to provide a STDM-based watermarking algorithm which exploits HVS properties related to the DWT coefficients. We design a pixel-wise masking vector that modulate the spreading vector in such a way that preserve its security. This paper is organized as follows, in section 2 a brief introduction on STDM will be presented. In section 3, the proposed approach is described. Then, in 3.1 we show where the watermark insertion occurs in the JPEG2000 decoding pipeline. In 3.2, we analyze the HVS properties related to DWT coefficients, and in 3.2 we present how we use these properties to adapt the STDM algorithm. Implementation and results are shown in section 4. Finally, we conclude in section 5.

Thanks to Belgian Interuniversity Attraction Pole IAP-VI for funding.

2. SPREAD TRANSFORM DITHER MODULATION

The basic idea of QIM, is the quantization of a signal sample x_i using a quantizer Δ_m chosen from a set of quantizers based on the embedding information m. For instance, to embed a message $m \in \{S1, S2\}$ in a host signal x, we need two different quantizers. If $m = S1$, we will use $Q_1 = Q(x_i, \Delta_1) = \lfloor (\frac{x_i}{\Delta_1}) \rceil \Delta_1$, where $\lfloor . \rceil$ denotes a rounding operation, and if $m = S2$, we will quantize the host signal sample x_i using $Q_2 = Q(x_i, \Delta_2) = \lfloor (\frac{x_i}{\Delta_2}) \rceil \Delta_2$. Detection in QIM is performed without access to the original data or the original watermark.

2.1. STDM

STDM is a special case of QIM. It exhibits robustness to re-quantization but maintains low complexity. Therefore, instead of quantizing the host signal x itself. The quantization occurs entirely in the projection of the host signal x onto a randomly generated vector p. Consequently, the embedding of each information bit occurs in the projection of the host signal x by quantizing it with a uniform, scalar, dithered quantizer [2]. Note that in this method, each bit of the watermark message is spread into *n* samples instead of one sample.

Let us consider a host signal $x = (x_1, x_2, ..., x_n, ...x_N)$ of length N. A random unit length vector p of length n is generated. For each n samples of the host signal x, one bit is spread in the projection of x onto a random vector p. Thus, the watermarked signal is given by:

$$y = x + (Q_m(x^T p) - x^T p)p \qquad m \in \{0, 1\} \tag{1}$$

The STDM decoder makes a decision based on the projection of the channel output y onto the spreading vector p. The detection can be then performed with a minimum distance decoder.

$$\hat{m} = arg \min_{\hat{m} \in \{0,1\}} |y^T p - Q_{\hat{m}}(y^T p)| \tag{2}$$

3. STDM-DWT ADAPTIVE WATERMARKING METHOD

In the proposed method, our aim is to adapt the STDM method in the DWT-domain, taking into account the fidelity vs. robustness trade-off while preserving the security of the watermark. Based on our previous work [6], we will model the vector p by using a pixel-wise vector masking. And instead of defining the vector p as a function from the local luminance of the image, we will generate a vector p that holds two major properties: security by preserving the randomness of p, and fidelity by exploiting perceptual properties related to HVS.

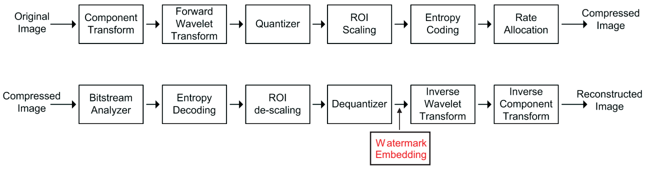

Fig. 1. JPEG2000 encoding and decoding pipeline

3.1. Exhibition Watermarking

In a Digital Cinema environment, the playback system should be able to insert the watermark identifying the time and location of the projection at the same rate when the motion pictures are decoded. In order to meet the real-time constraint, the complexity of the watermarking algorithm should be as low as possible.

In the JPEG2000 decoder, the image is already decomposed into DWT coefficients and the inverse DWT will be performed in the following step of the decoding framework. Hence, the embedding process consists uniquely in using the technique described in the following. Some pre-processing on the de-quantized wavelet coefficients will be done in order to improve the system performance. The proposed watermarking method is applied in the DWT-domain at the decoding phase. This method is designed to fit in the JPEG2000 decoding pipeline as shown in figure 1. Embedding is done after entropy decoding and de-quantization. Thus, that context enable us to work directly on wavelet coefficients in their right location.

3.2. DWT perceptual modeling

The DWT properties linked to HVS are generally exploited in three fields [3] as described in the following:

- The Human eye is less sensitive to the noise in high resolution bands, and especially in the diagonal bands.
- The changes in high or low brightness areas of the image are less perceptible.
- The Human eye is less sensitive to noise in highly textured areas but, among these, more sensitive near the edges.

In the proposed method, only the luminance component is marked. This choice is mainly motivated by robustness reasons. Note that, JPEG compression attack preserves better the luminance component than the chrominance components. In addition, the human eye is more sensitive to slight color changes than to slight luminance changes.

3.3. Adaptive and Secure STDM

In STDM, the quantity by which each DWT coefficient can be altered, is initially controlled by the quantization error. From (1) we can see that change occurs entirely in the direction of the spreading random vector p. We propose the inclusion of a perceptual model in the STDM framework by modeling the vector p in the direction of least perceptual distortion.

We will define a pixel-wise masking vector v based on the three fields described in the previous section, that will modulate the random vector p. Consequently, the modeled vector will ne denoted p' such that:

$$[p'] = [v][p] \tag{3}$$

In this way, we can overcome the security drawback of choosing a vector p that is a function of the host signal. Hence, we preserve the security of the vector p by holding its randomness and at the same time modeling it by a pixel-wise vector that considers the properties of the HVS. Therefore, each entry of the vector v evaluates the amount by which each DWT coefficient can be modified. It corresponds to a weighting function that is the product of three main constraints: Resolution, Luminance, Texture and Edges.

Resolution

The first constraint is used to define in which decomposition level the watermark embedding occurs. As DWT has poor directional selectivity for diagonal features (there is only one filter for diagonal features), no embedding is done in diagonal sub-bands. Initially, we decided to embed information bits in the third-level horizontal and vertical sub-bands $(LH3, HL3)$ of a $k = 5 - level$ wavelet decomposition. We assume that in 3^{rd} level decomposition $l = 3$, more detail coefficients are available than in previous decomposition levels and are less vulnerable than those in the following levels.

Luminance

To indicate the capacity of each DWT coefficient to support changes based on Luminance perceptual properties, we compute the local luminance in the embedding resolution based on the gray level values of the low pass version of the image as follows:

$$L(l,i,j) = (\frac{1}{I^{LL}_{kmax}})I^{LL}_{k}(1 + \lfloor\frac{i}{2^{k-l}}\rfloor, 1 + \lfloor\frac{j}{2^{k-l}}\rfloor) \tag{4}$$

As the human eye is less sensitive in very dark and bright regions, we intend to introduce more weight during the embedding process of the corresponding points. Consequently, the local luminance factor is modified as follows:

$$L'(3,i,j) = \left\{ \begin{array}{ll} 1 - L(3,i,j) & if L(3,i,j) < 0.5 \\ L(3,i,j) & \text{otherwise} \end{array} \right\} \tag{5}$$

Texture and Edges

The discrete wavelet transform performs good decorrelation of the signal and distributes the energy of the image on a few significant coefficients. It also provides a good tool for edges and textures detection. However, high activity regions are essential for both compression and watermarking: compression needs to preserve their integrity and the HVS is less sensible to modifications in these regions, which is a useful property for watermarking. In the masking model proposed in [3], edges and textures activities need the analysis of several sub-bands, which may increase computational complexity of the watermarking method. Instead of that and for computational efficiency, we use the diagonal sub-band $HH3$ as a significant index for edges and textures details.

Hence, each entry of the vector v will be generated as follows:

$$\begin{bmatrix} v(1) \\ v(2) \\ . \\ . \\ v(n) \end{bmatrix} = \begin{bmatrix} L'_3(1) \\ L'_3(2) \\ . \\ . \\ L'_3(n) \end{bmatrix} \begin{bmatrix} \epsilon & + & HH_3(1) \\ \epsilon & + & HH_3(2) \\ . & . & . \\ . & . & . \\ \epsilon & + & HH_3(n) \end{bmatrix}^T \tag{6}$$

ϵ is a constant parameter introduced to avoid the neutralization of the luminance constraint in case of insignificance texture and edges details in the HH3 sub-band. ϵ should be small and will be defined in the implementation.

Another alternative to exploit edges and textures sensitivity, is to apply a non-linear scaling function before the embedding. Thus, applying the corresponding inverse scaling after embedding. And this to put more watermark energy in edges and textured areas. Hence, the host vectors after scaling the coefficients can be written as follows:

$$x(i,j) = sign(h(i,j)) \cdot |h(i,j)|^{\beta}, \qquad \beta \leq 1 \tag{7}$$

We will analyze the performance of the proposed scheme with the non-linear scaling function or without it. Note that, this step can introduce some distortions. The scaling factor β is controlled in a way to avoid any perceptual distortions. To embed the watermark, a non-overlapping window of size $(n \times 1)$ is slid over the $HL3$ and $LH3$ coefficients. The size of the window is determined by the coding rate and the size of watermark to embed. At each window position one bit of the watermark is embedded by using the adaptive Spread-Transform Dither Modulation scheme introduced in section 2.1.

4. IMPLEMENTATION AND RESULTS

A set of eight 2K-raw frames of digital cinema content were obtained and are subjected to the watermark insertion process. Test material was obtained from StEM DCI material [7], RED cameras sample shots and Big Buck Bunny open movie project. All images have a 1920x1080 pixel spatial resolution and are 8-bit depth for each color component channel.

As specified for JPEG2000 compression of 2K resolution content in [7], a 5-level $9/7$ wavelet decomposition is performed firstly. The total information payload is $35bits$, embedded in the level-3 of $LH3$ and $HL3$ entire sub-bands. The watermark embedding strength is controlled by the quantization step size Δ used in STDM. We take $\epsilon = 1$ and the scaling factor parameter β is fixed to 0.85. Experiments were performed to determine both fidelity and robustness of the watermarking algorithm.

4.1. Perceptual Quality

Subjective and quantitative evaluation of the distortion introduced due the watermarking process has been analyzed. As mentioned previously, the watermark robustness depends directly on the embedding strength, which in turn influences the visual degradation on the image. In the proposed scheme, we fix the embedding intensity Δ, however this quantization step is applied differently at each coefficient due to the masking vector. The number of the embedded data is 70 bits/frame and PSNR stands for the average value for several repetitions on each image.

Butterfly tests on large-screen projections were performed with five critical observers to evaluate image fidelity. Each of the eight test frames were watermarked with three different embedding intensities, keeping always the PSNR above 55 dB. Watermarked frames were presented in couple respectively with the original one, and observers were required to detect the watermarked frame in each couple. The results show that no perceptible distortion is produced in the watermarked images for PSNRs above 55 dB. In any case, none of the viewers was able to reliably identify the watermarked frame. A sample watermarked frame is shown in figure 2.

4.2. Robustness and Security

We have considered two possible attack scenarios: Perfect Capture and Camcorder Capture. In Perfect Capture, we assume that the decoded bit-stream can be captured from some digital output port after embedding without any distortions. Experimental results show that the proposed method is robust against image processing attacks. A 5% quality image compression rate degrades severely the image so to consider the algorithm robust to compression. Figure 3 shows that the proposed method gives lower BER compared to traditional STDM and other STDM improved methods. The STDM-Lum curve illustrates the results when the projection vector is defined from the local luminance of each pixel. And STDM-Lum Beta=0.85 when we apply a non-linear scaling function with a scaling factor $\beta = 0.85$. And finally, STDM-SecAdap illustrates the results for the proposed method with or without a non-linear scaling.

One of the major properties of STDM, is the fact of having a security protection which improves the watermarking system security. In the proposed method, the security level of the STDM is preserved by using a secret direction which is modulated by a pixel-wise masking vector.

4.3. Anti-camcording

As can be seen in the sample frame in figure 4, the frame capture introduce arbitrary scaling, rotation, some perspective projection, blurring and compression. Each captured frame has been processed such that all black-bars or boundary pixels are removed, resulting on approximately (500x375) resolution images. Detection is performed after individual image registration using the morphon method [8]. Registered images present a PSNR of about 27-28dB.

Fig. 2. From Big Buck Bunny film - DCI

5. CONCLUSION

We have proposed a robust and secure exhibition watermarking algorithm. The embedding occurs in the decoding pipeline before the projection of a Digital Cinema movies. The goal is to deter capture and unauthorized redistribution of electronically distributed movies. The algorithm has a very low complexity and is suitable for implementation inside the JPEG2000 decoder. The tests show that the proposed method highly preserves the image quality. The watermark survives low bit-rate compression and capture with a video camera, allowing us to reach the watermark recovering constraint of the DCI without compromising image quality.

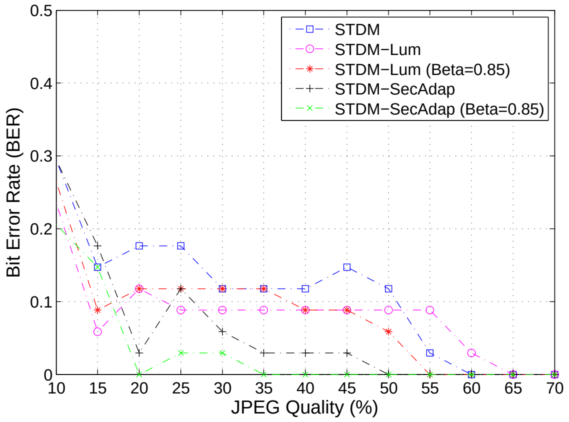

Fig. 3. Bit error rate (BER) vs. JPEG compression using an embedding rate of 1/120 and a PSNR of around 60dB

6. ACKNOWLEDGEMENTS

The work of Rony Darazi is supported by BCRYPT project, a Belgian Interuniversity Attraction Pole IAP-VI fund programme.

7. REFERENCES

[1] I.J. Cox, J. Kilian, F.T. Leighton, and T. Shamoon, “Secure spread spectrum watermarking for multimedia,” *Image Processing, IEEE Transactions on*, vol. 6, no. 12, pp. 1673–1687, Dec 1997.

[2] B. Chen and G.W. Wornell, “Quantization index modulation: a class of provably good methods for digital watermarking and information embedding,” *Information Theory, IEEE Transactions on*, vol. 47, no. 4, pp. 1423–1443, May 2001.

[3] M. Barni, F. Bartolini, and A. Piva, “Improved wavelet-based watermarking through pixel-wise masking,” *Image Processing, IEEE Transactions on*, vol. 10, no. 5, pp. 783–791, May 2001.

[4] Qiao Li and I.J. Cox, “Using perceptual models to improve fidelity and provide resistance to valumetric scaling for quantization index modulation watermarking,” *Information Forensics and Security, IEEE Transactions on*, vol. 2, no. 2, pp. 127–139, June 2007.

[5] Dong Yu, Lihong Ma, Guoxi Wang, and Hanqing Lu, “Adaptive spread-transform dither modulation using an improved luminance-masked threshold,” Oct. 2008, pp. 449–452.

[6] P. Callau, R. Darazi, and B. Macq, “Exhibition QIM-based Watermarking for Digital Cinema,” in *Media Forensics and Security*. in Proc. 2009, SPIE.

[7] “http://www.dcimovies.com/,” .

[8] H. Knutsson and M. Andersson, “Morphons: segmentation using elastic canvas and paint on priors,” Sept. 2005, vol. 2, pp. II–1226–9.

(a) 2k-Shot (b)Cropped

(c)Adjusted

Fig. 4. Pre-processing for Watermark reconstruction