ADAPTIVE WATERMARKING IN THE DCT DOMAIN

Bo Tao *Bradley Dickinson*

Department of Electrical Engineering
Princeton University, Princeton, NJ 08544
Email: botao, bradley@ee.princeton.edu

ABSTRACT

An adaptive watermarking technique is introduced in this work. A regional perceptual classifier is employed to assign a noise sensitivity index to each region. The watermark is inserted in the original image according to this index by using block DCT. The detection of the watermark is designed to achieve a desired false alarm probability.

1. INTRODUCTION

With advances in computer network and multimedia technology, digital media is rapidly proliferating, calling for greater copyright protection for electronic publishing. Digital watermarking is a technique to protect intellectual property in digital form. By adding an invisible signal (called watermark) to the original media, it is hoped that such a signal will provide evidence of the legal ownership or at least help the owner to detect copyright violations [3].

Many watermarking techniques [1] have been proposed, mainly focusing on the invisibility of the watermark and its robustness against various signal manipulations and hostile attacks. The first group of techniques work in the spatial domain, for example, changing the LSB of some pixels [8], recording the difference between randomly selected pairs of points [1], and so on. These techniques can suffer from signal compression and hostile attacks. Another group of techniques work in the spatial-frequency domain and add a watermark by manipulating various frequency elements [2]. Frequency domain techniques are much more robust against compression and geometrical transformations than spatial domain techniques. Nevertheless, partly inspired by [5], we notice that one weakness for many previous spatial frequency domain approaches is that the human visual system (HVS) is not taken into account when selecting positions to insert the watermark. Because of the invisibility constraint of a watermark, these techniques have to use signals of relatively lower power than would otherwise be possible, to avoid degrading the image quality, inevitably limiting the robustness of the watermark. Among forthcoming works, reference [6] has been mentioned to us as being of some relevance.

We make an important observation in this work: a good watermarking technique has to adapt to the particular image being watermarked in order to exploit specific HVS characteristics and hence embed a strong signal. We propose an adaptive watermarking technique, which assigns each spatial region a noise sensitivity label and embed the watermark using block DCT according to this sensitivity label. The watermark detection threshold is chosen to achieve a desired false alarm probability, which we believe is an appropriate performance measure. The adaptive watermarking technique and its detection will be discussed in section 2, with experimental results given in section 3. The final section summarizes the paper.

2. ADAPTIVE WATERMARKING AND ITS DETECTION

When adding a watermark to an image, several questions can be asked: where to add the watermark, how much energy can be inserted, and how well can it be retrieved? These questions are not isolated and our technique can be viewed as an attempt to answer these three questions simultaneously. The system diagram is given in Fig. 1.

2.1. DCT domain watermarking

We choose to work in the block DCT domain for the following reasons:

- DCT has good energy compaction capability,
- it is feasible to incorporate the HVS characteristics in this domain, and

[1] We concentrate on watermarking still images in this work, although similar principles can be applied to audio and video signals.

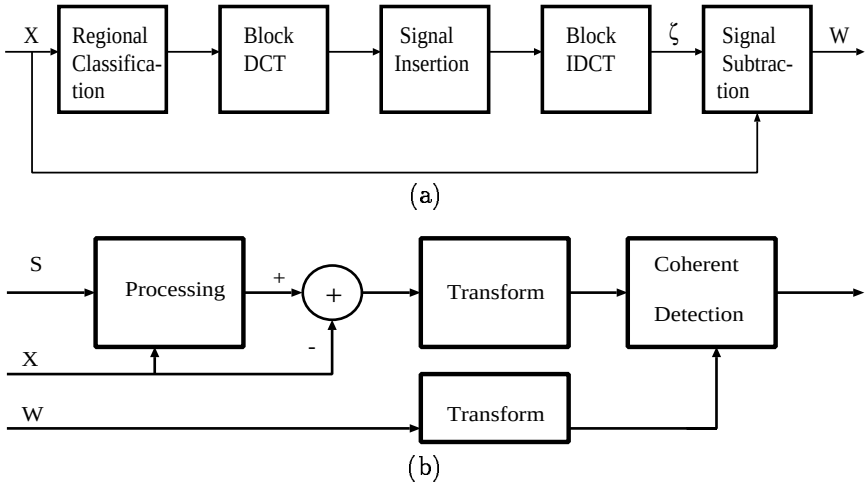

Figure 1. Watermarking system diagram. (a): watermarking; (b): watermark detection. X, W and S are the original image, the watermark and the suspected image, respectively.

- the sensitivity of HVS to the DCT basis images has been extensively studied [4] resulting in a default JPEG quantization table [12].

Generally speaking, the watermark has to be added to frequencies of high energy in order to be resistant to noise. In our system, the N AC coefficients having the smallest quantization step sizes in the aforementioned table are selected for this purpose. Ideally, one may think that the largest perturbation to the ith coefficient in a block is given by $D_i/2$, where D_i is the ith quantization step size in the JPEG table. However, in our experiments, we find this leads to visible artifacts when many coefficients are perturbed in this fashion. Instead, we propose the following formula to add a signal at the ith AC position in a block,

$$\hat{\xi}(i) = \hat{x}(i) + \max[\hat{x}(i) \times \alpha_m,\ sgn(\hat{x}(i)) \times D_i/\kappa] \quad (1)$$

where m is the noise sensitivity index for the block, $\hat{x}(i)$ is the ith DCT coefficient with $\hat{\xi}(i)$ being its watermarked counterpart, and κ satisfies $5 \leq \kappa \leq 6$. Notice that the watermark signal is not generated randomly as in [2]. This is because randomized watermarks can be defeated by obtaining multiple copies and "averaging out" the randomness [9].

As indicated in the above formula, a noise sensitivity index is assigned to each block. This can effectively exploit various masking effects of the HVS, both to protect regions (such as edges) sensitive to noise and insert strong signals in regions with low noise sensitivity. The regional classification algorithm is introduced in the following section.

2.2. Spatial adaptive placement of the watermark

We realize the similarity between watermarking and quantization, and propose to classify each block into different noise sensitivity classes and insert signals of different energies accordingly (eq. (1)).

In our classification algorithm [10], such properties as luminance masking, edge masking and texture masking effects of the HVS are exploited. It classifies a block into one of six perceptual classes, from 1 to 6, edge, uniform with moderate intensity, uniform with either high or low intensity, moderately busy, busy and very busy, in descending order of noise sensitivity.

Let the average gradient magnitude of the picture be denoted by G. A block is an edge block *only if* the number of pixels satisfying

$$mag(\nabla x) > k \times G$$

is greater than K, with k and K being pre-selected constants. However, such a simple counting alone will inevitably result in many textured blocks being misclassified as edge blocks, due to their high spatial variations. To solve this problem, a second test is used to decide whether a block having passed the counting test is an edge or textured block. Let the variances of the 8-connected neighboring blocks of the current block be σ_i, $i = 1, ...8$, numbered in counter clockwise order beginning from its east neighbor. A block containing a vertical edge passing between a uniform ares and a textured area will satisfy

$$\begin{aligned} 2 \quad \leq \quad & I(\sigma_2 > r \times \sigma_4) + I(\sigma_4 > r \times \sigma_2) + \\ & I(\sigma_1 > r \times \sigma_5) + I(\sigma_5 > r \times \sigma_1) + \\ & I(\sigma_6 > r \times \sigma_6) + I(\sigma_6 > r \times \sigma_8) \end{aligned}$$

where $I(*)$ is the indicator function and r is a constant. Similar tests can be set up for horizontal, diagonal and anti-diagonal edges and for the situation where edges passing between two uniform areas. We call them ratio tests. So, a block is declared an edge block if it satisfies the previous counting test and at least one of the ratio tests.

For a non-edge block, its variance will be used to further classify it into class $3-6$, with class 3 containing blocks with small variance. Class 3 is further divided into two classes, 2 and 3, with class 2 having moderate intensity. The reason for differentiating class 2 and 3 is that relatively uniform regions with moderate intensity are more sensitive to noise than regions with similar variance, but either high or low intensity [11].

2.3. Watermark detection

The watermark detection process is given in Fig. 1(b), where the suspected image w may go through certain processing such as registration and padding if necessary, mainly to take care of geometrical transformations. For the same reason, the detection is performed in the spatial-frequency domain, using either DFT or DCT (but *NOT* block transform). Let $y = s - x$, and $\hat{y}$ and $\hat{w}$ be the transform of y and w, respectively. The detection is naturally modeled as a hypothesis testing problem,

$$H_0 \quad : \quad \hat{y}_i = n_i$$

vs

$$H_1 \quad : \quad \hat{y}_i = \hat{w}_i + n_i$$

where n_i is the noise in the ith frequency position.

By assuming $\{n_i\}$ are *i.i.d.* Gaussian with zero mean and variance σ^2, the critical region Γ is given by

$$T(\hat{y}) = \frac{\sum_{i=1}^{M} \hat{y}_i \hat{w}_i}{\sigma \times \sqrt{\sum_{i=1}^{M} \hat{w}_i^2}} > \gamma \qquad (2)$$

where M is the number of low-pass frequency positions used for detection. Then the false alarm probability and detection probability are given by

$$P_0(\Gamma) = 1 - \Phi(\gamma) \qquad (3)$$

and

$$P_1(\Gamma) = 1 - \Phi(\gamma - \frac{1}{\sigma}\sqrt{\sum_{i=1}^{M} \hat{w}_i^2}) \qquad (4)$$

respectively, where Φ is the cumulative probability distribution function for an $N(0,1)$ random variable.

We believe that a low false alarm probability is important for a watermark to be practically useful. Therefore, during the detection process, the critical region is found by constraining $P_0(\Gamma) \leq \beta$ and using eq. (3), with β being the desired false alarm probability. In this way, one gets a family of receiver operating characteristics (ROC) by parameterizing on σ, and can choose a decision point by estimating the noise power. The above formulation is optimal (under the assumptions above) according to the Neyman-Pearson theorem [7].

3. EXPERIMENTAL RESULTS

The above watermarking technique has been tested on various images. Here we give the results using the image Lena. The parameters are so chosen that $N = 11$, $M = 500$, $k = 4.0$, $K = 4$, and α_i are 0.1, 0.12, 0.15, 0.2, 0.25 and 0.3 from class 1 to 6. The original image, the watermarked image and the classification result are given in Fig. 2.

A total of 250 images are used for detection test. Among them, images 1-6 are JPEG compressed Lena with compression ratios ranging from 1:7 to 1:43 (with quality from 50% to 5% correspondingly). Images 7-126 are the results of adding uniformly distributed random noise to Lena, with PSNR ranging from 22.1 to 29.4dB. The next 120 images, numbered 127-246, are watermarked Lena corrupted with uniformly distributed noise with the same PSNR range. The last 4 images are JPEG compressed watermarked Lena with compression ratios between 1:11 and 1:43. The detection statistics (eq. (2), assuming $\sigma = 1$) are given in Fig. 3. The *maximum* and average detection statistics for the first 126 non-watermarked images are 9.92 and -0.36 respectively, while the *minimum* and average values for the next 124 watermarked images are 685.77 and 771.19, respectively.

4. CONCLUSION

An adaptive watermarking technique is proposed in this paper, which explicitly exploits HVS characteristics to protect regions vulnerable to noise and insert watermarks of larger energy in regions with higher perceptual masking ability through the use of a regional classifier. The Neyman-Pearson theorem is used to design the detector to achieve a desired low false alarm probability, which is arguably important in practice.

To a large extent, watermarking is equivalent to quantization and, hence many existing techniques in the area of perceptual coding can be exploited. For example, a just noticeable distortion (JND) profile estimator can be very useful in determining the maximum amount of energy that can be inserted without causing visual artifacts. On the other hand, a space-frequency decom-

(a) (b) (c)

Figure 2. (a): the original image; (b): the watermarked image; (c): the perceptual classification result. In (c), regions with the highest noise sensitivity are the darkest, while regions with the lowest noise sensitivity are the brightest.

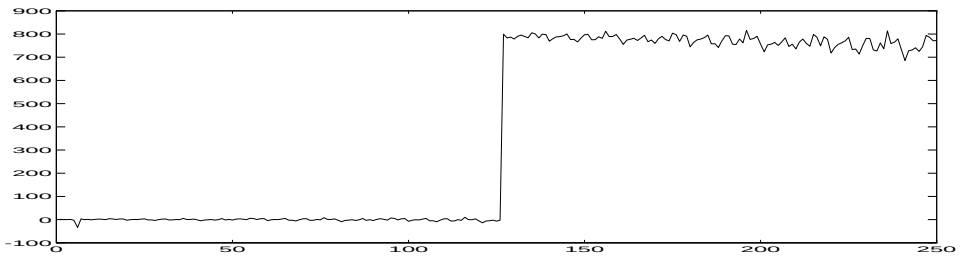

Figure 3. Detection statistic comparison. Images 1-126 are non-watermarked, while images 127-250 are watermarked.

position such as wavelet transform may provide a domain where both classification and signal insertion can be done in a systematic way, avoiding going back and forth between the space and frequency domains.

References

[1] W. Bender, D. Gruhl and N. Morimoto, "Techniques for Data Hiding", *SPIE vol. 2420,* 1995.

[2] I. Cox, J. Kilian, T. Leighton and T. Shamoon, "Secure Spread Spectrum Watermarking for Multimedia", *NEC Research Inst. Tech. Report, 95-10,* 1995.

[3] S. Craver, N. Memon, B. Yeo and M. Yeung, "Can Invisible Watermarks Resolve Rightful Ownerships?", *IBM Research Report, RC 20509*, 1996.

[4] H. Lohscheller, "A Subjectively Adapted Image Communication System", *IEEE Trans. Communications, vol. 32, no. 12,* 1984.

[5] C. Podilchuk, *personal communication.*

[6] C. Podilchuk and W. Zeng, "Digital Image Watermarking using Visual Models", *to appear in Human Vision and Electronic Imaging II, SPIE*, 1997.

[7] H. V. Poor, "An Introduction to Signal Detection and Estimation", *second edition, Springer-Verlag,* 1994.

[8] R. Schyndel, A. Trikel and C. Osborne, "A Digital Watermark", *Proc. ICIP,* 1994.

[9] H. Stone, "Analysis of Attacks on Image Watermarks with Randomized Coefficients", *NEC Research Inst. Tech. Report*, 1996.

[10] B. Tao and B. Dickinson, "Adaptive Bit Allocation with Applications to MPEG Video Coding", *to be submitted to IEEE Trans. Cir. Sys. Video Tech.*.

[11] B. Tao, "On Adaptive Quantization and Rate Control for MPEG Video Coding Environments", *David Sarnoff Technical Report,* 1996.

[12] G. Wallace, "The JPEG Still Picture Compression Standard", *IEEE Trans. Consumer Electronics, vol. 38, no. 1,* 1992.