Optimum Detection for Spread-Spectrum Watermarking that Employs Self-masking

Wei Liu, Lina Dong and Wenjun Zeng

Computer Science Department, University of Missouri-Columbia, Columbia MO 65211, USA

ABSTRACT

Digital watermarking is an effective and promising approach to protect intellectual property rights of digital media. Spread spectrum (SS) is one of the most widely used image watermarking schemes. In SS watermarking, the watermark signal is usually modulated by the just-noticeable difference (JND) of the host image. The JND is measured by advanced perceptual models as a non-linear function of local image features. In this paper, the optimum detection scheme for such non-linearly embedded watermarks is addressed. Closed-form detectors are found for arbitrary JND models that exploit the self-masking property of the human visual system.

Index Terms—digital watermarking, watermark detection, spread spectrum, human visual system, perceptual masking, Neyman-Pearson criterion

1. INTRODUCTION

Digital watermarking has been demonstrated to be effective in protecting intellectual property rights. It plays an important role in multimedia security applications. In this paper, we focus on watermarking of images, although the general idea can be easily extended to other forms of media.

In designing image watermarking schemes, robustness and invisibility are two desirable features. There is always a tradeoff between these two requirements. Among others, spread-spectrum (SS) watermarking [1] is an important approach because of its robustness to attacks and easiness to exploit properties of the human visual system (HVS). To achieve the best invisibility-robustness tradeoff, the watermark signal in an SS system is usually modulated by the just-noticeable-difference (JND). Now the questions to be answered are 1) how to model the JND and 2) how to optimize the detector based on the JND model.

In the literature, additive/multiplicative watermarking schemes are widely used approaches [2]. The former uses a small constant to control the watermark strength while the latter modulates the watermark by the amplitude of the local image feature. Both implicitly employ a linear perceptual model, i.e., the JND is modeled as a linear function of the local image feature, which is not precise enough to exploit the HVS properties. In comparison, Podilchuk and Zeng's work [3] is more sophisticated. They employ a *non-linear* self-masking model to measure the perceptual redundancy, which is based more closely on the research results of human vision, and has shown better visual quality than the original SS watermarking scheme [3].

As for the detection schemes of SS watermarking, the linear correlation detector (LCD) has been extensively used. However, it has been proved that LCD is not an optimum solution unless the host signal is white Gaussian and the embedding is additive. In recent years, advanced researches on optimum detectors have been reported. Readers are referred to [4][5][6][7] for detectors designed for additive watermarks, and [8][9] for multiplicative ones. However, no optimum detection scheme for watermarks based on *non-linear* perceptual models have been reported except for the authors' preliminary work [10], in which the watermarks are embedded using Podilchuk and Zeng's approach, and the host signal is assumed to follow the generalized Gaussian distribution (GGD). In this paper, we extend our study to *arbitrary* JND models that exploit the self-masking property of the HVS, with arbitrary host-signal distributions.

The rest of the paper is organized as follows. Section 2 formulates the non-linear embedding process that exploits the self-masking property of the HVS. The optimum detection is derived in Section 3. Simulation results are presented in Section 4. Section 5 concludes the paper.

2. PERCEPTUAL SPREAD-SPECTRUM WATERMARKING WITH SELF-MASKING

The block-diagram of perceptual SS watermarking is illustrated in Fig. 1. Let $\mathbf{x} = \{x_1, x_2, \ldots, x_N\}$ be a set of image features to be watermarked. In the majority of state-of-the-art watermarking schemes, the features are selected to be coefficients in a particular transform domain such as the DCT, DWT or DFT domain, to efficiently exploit the HVS properties and to be compliant with image compression standards. We assume x_i's are i.i.d. random variables, of which the probability density function (PDF) is f_X. Let $\mathbf{w} = \{w_1, w_2, \ldots, w_N\}$ be the watermark signal of which the elements are pseudo-randomly generated based on a secret key. Without loss of generality, let $E\{w_k\} = 0$ and $D\{w_k\} = 1$ for $k = 1, 2, \ldots, N$.

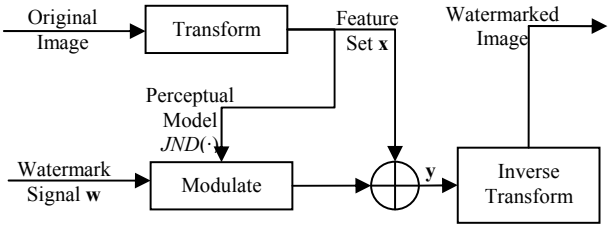


Fig. 1: Block-diagram of perceptual SS watermark-embedding. The watermark signal is first modulated by the JND before embedding

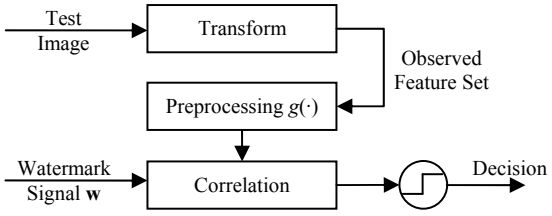


Fig. 2: Block-diagram of a generalized correlation detector. The observed features are pre-processed before the correlation.

Sophisticated embedding strategy employs the HVS properties and maximizes the embedding strength to the JND of the image by:

$$y_k = x_k + JND_k \times w_k, (k = 1, 2, \ldots, N) \quad (1)$$

where JND_k is the largest possible distortion to x_k without being noticed by human eyes. One possible setting is to let w_k take the value from $\{+1, -1\}$ to fully exploit the HVS.

Although JND_k is in general related to x_j even if $j \neq k$, HVS studies reveal that x_k plays the most important role in JND_k. In other words, the distortion to x_k introduced by the watermark is best concealed by x_k itself. This phenomenon is also known as *self-masking*. Most watermarking schemes focus on self-masking. In this paper, we restrict JND_k to be a single variable function of x_k with $JND_k = JND(x_k)$, and formulate the perceptual SS watermark embedding as

$$y_k = G(x_k) = x_k + JND(x_k) \times w_k, (k = 1, 2, \ldots, N). \quad (2)$$

3. THE OPTIMUM DETECTION SCHEME

Given any observed sample y, the detector needs to make a decision between two hypotheses:

$$\begin{aligned} H_1 &: y = x + \theta \times JND(x) \times w \\ H_0 &: y = x \end{aligned}. \quad (3)$$

Here we introduce a factor θ to denote the strength of the watermark. If the watermark is not supposed to be attacked before detection, $\theta = 1$. But in most cases we are more interested in detecting watermarks in degraded images, where the watermark strength is decreased to a small positive number. This is also known as robust detection [8].

In our previous work [10], for a particular JND model [13] we propose to transform the test features to a perceptually uniform domain, where the non-linearly embedded watermarks turn out to be additive ones. Then the Bayesian hypothesis testing can be applied to derive a closed-form solution. The resulting locally optimum detector (LOD) is a generalized correlation detector (GCD), of which the block diagram is shown in Fig. 2. A GCD pre-processes each observed feature point before the correlation with the watermark signal, i.e.

$$GCD(\mathbf{y}) = \frac{1}{N}\sum_{k=1}^{N} g(y_k) w_k \,. \quad (4)$$

In this section, we do not follow the perceptually-uniform approach. In stead, we will explicitly find an optimum pre-processing $\hat{g}(x)$ for a given JND model, which achieves the best error performance among all GCD's according to the Neyman-Pearson criterion.

3.1. Error Performance of a GCD

According to the central limit theory (CLT), the output of a GCD is a Gaussian random variable. More specifically, for a particular image feature set **y**, if not watermarked, a GCD will output a Gaussian variable with the mean m_0 and the standard deviation σ_0; if it is watermarked, the distribution will have the mean and the standard deviation being m_1 and σ_1. It is the 4-tuple (m_0, σ_0, m_1, σ_1) that determines the error performance of a GCD.

To detect the existence of a watermark, the decision is made by comparing the GCD's output to a pre-defined threshold. There are two types of errors that could occur: an error of false alarm occurs if the output is greater than the threshold but the image is actually not watermarked; on the other hand, if the output is smaller than the threshold for a watermarked image, an error of miss occurs. Generally, the decision threshold η is chosen such that the probability of false alarm (P_f) is fixed. Then, an optimum detector should minimize the probability of miss (P_m). In other words, the probability of detection ($P_d = 1 - P_m$) should be maximized. This is also known as the Neyman-Pearson criterion [11].

By assuming that the output of a GCD is asymptotically Gaussian, we can calculate P_f and P_d as follows

$$P_f = Q\left(\frac{\eta - m_0}{\sigma_0}\right), \; P_d = Q\left(\frac{\eta - m_1}{\sigma_1}\right). \quad (5)$$

where $Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^{+\infty} \exp\left(-t^2/2\right) dt$.

The receiver operating characteristics (ROC) describes the relationship between P_f and P_d:

$$P_d = Q\left(\frac{Q^{-1}(P_f) \times \sigma_0 + m_0 - m_1}{\sigma_1}\right). \quad (6)$$

Now we will estimate the 4-tuple (m_0, σ_0, m_1, σ_1) for a GCD. Under hypothesis H_0, the mean of (4) is

$$m_0 = E_{\mathbf{w}}\{GCD(\mathbf{y}) | H_0\} = E_{\mathbf{w}}\left\{\frac{1}{N}\sum_{k=1}^{N} g(x_k) w_k\right\} = 0, \quad (7)$$

and the variance is

$$\sigma_0^2 = E_{\mathbf{w}}\left\{(GCD(\mathbf{y}) - m_0)^2 \middle| H_0\right\} = \frac{1}{N^2}\sum_{k=1}^{N} [g(x_k)]^2 \,. \quad (8)$$

Under hypothesis H_1, the mean of (4) is

$$m_1 = E_{\mathbf{w}}\left\{GCD(\mathbf{y})\middle|H_1\right\} = E_{\mathbf{w}}\left\{\frac{1}{N}\sum_{k=1}^{N} g(x_k + \theta \times JND(x_k)w_k)w_k\right\}$$
$$\approx E_{\mathbf{w}}\left\{\frac{1}{N}\sum_{k=1}^{N}\left[g(x_k) + g'(x_k)\times\theta\times JND(x_k)w_k\right]w_k\right\} \quad (9)$$
$$= \frac{\theta}{N}\sum_{k=1}^{N} g'(x_k)JND(x_k)$$

where the approximation is based on the first-order Taylor-series, assuming the watermark can be considered as a weak signal. There is an implicit assumption that $g(x)$ is derivable. Similarly we derive that $\sigma_1^2 \approx \sigma_0^2$ for weak watermark signals. Now the ROC in (6) can be simplified as

$$P_d \approx Q\left(Q^{-1}(P_f) - \frac{m_1}{\sigma_1}\right). \quad (10)$$

3.2. The Optimum GCD

The error performance of a GCD can be measured by m_1/σ_1 since $Q(x)$ is monotonically decreasing. Thus the optimum GCD should have the pre-processing $\hat{g}(x)$ maximizing:

$$\frac{m_1}{\sigma_1} = \frac{\frac{\theta}{N}\sum_{k=1}^{N} g'(x_k)JND(x_k)}{\sqrt{\frac{1}{N^2}\sum_{k=1}^{N}\left[g(x_k)\right]^2}}. \quad (11)$$

If we roughly assume **x** is an ergodic process, the time average in (11) can be replaced with the ensemble average:

$$h(g(x)) = \frac{\int_{-\infty}^{\infty} g'(x)JND(x)f_X(x)dx}{\sqrt{\int_{-\infty}^{\infty}\left[g(x)\right]^2 f_X(x)dx}}. \quad (12)$$

By noticing that $h(g(x)) = h(\rho \cdot g(x))$ when ρ is a positive constant number, we can always normalize the denominator and assume that

$$\int_{-\infty}^{\infty}\left[g(x)\right]^2 f_X(x)dx = 1 \quad (13)$$

and equivalently, we are to find the optimum $\hat{g}(x)$ which maximizes the *numerator* of (12) conditioned on (13). This is usually solved by exploiting a Lagrange multiplier

$$\hbar\left[g(x)\right] = \left[\int_{-\infty}^{\infty} g'(x)JND(x)f_X(x)dx\right]^2 - \mu\int_{-\infty}^{\infty}\left[g(x)\right]^2 f_X(x)dx \quad (14)$$

We solve (14) by using the Euler-Lagrange differential equation [14] in calculus of variations and derive the optimum pre-processing to be

$$\hat{g}(x) = -\frac{\left[f_X(x)JND(x)\right]'}{f_X(x)} \quad (15)$$

So far we have found the optimum GCD which provides the best error performance based on the Neyman-Pearson criterion. This result, not surprisingly, agrees with the LOD derived from the perceptually-uniform domain. More detailed derivations can be found in [12].

4. SIMULATION RESULTS

In this section, we use the DCT-domain watermarking approach proposed in [3] as an example to test the performance of our detector. A comparison is made with the LCD, which does not perform any pre-processing before correlation (that means $g(x) = x$ in (4)).

The embedding procedure of [3] is summarized as follows: the host image is decomposed using 8×8 block DCT, resulting in one DC subband and 63 AC subbands. For each AC coefficient x_k, the JND for x_k is calculated using a non-linear transducer [13]:

$$JND(x) = \max\left\{C_{T0}, C_{T0}\left|\frac{x}{C_{T0}}\right|^{\varepsilon}\right\}. \quad (16)$$

where C_{T0} is the basic threshold (see [3] for details). If x_k's amplitude is greater than $JND(x_k)$, watermark is embedded using (2); otherwise it is left unchanged. The reason to embed watermarks only to those coefficients with large amplitudes is that the weak features are vulnerable to attacks such as quantization. In this case, it is easily proved that the following non-linearity always holds for any image feature that is chosen for watermark embedding

$$JND(x) = C_{T0}\left|\frac{x}{C_{T0}}\right|^{\varepsilon}. \quad (17)$$

For DCT/DWT domain AC coefficients, a commonly used model is the GGD with zero mean. That is

$$f_X(x) = A\ \exp(-|\beta x|^c) \quad (18)$$

where A, β and c can be estimated based on the test image.

The LOD is derived from (3) and (15) as

$$LOD(\mathbf{y}) = \frac{1}{N}\sum_{k=1}^{N} sign(y_k)\left|\frac{C_{T0}(k)}{y_k}\right|^{1-\varepsilon} \times\left(c_k\left|\beta_k y_k\right|^{c_k} - \varepsilon\right)\times w_k \quad (19)$$

where c_k and β_k depends on the subband of y_k. In the following simulation, we assume the AC coefficients of an image are roughly i.i.d. Laplacian (c_k = 1). This is a simplified approach, but still works well in the simulations.

We first compare the detection performances of the LOD and the LCD on watermarked images without attacks. Three 512×512 Lena images is used for testing. Both of the two detectors are normalized such that their outputs have unit variance under hypothesis H_0 (σ_0=1), thus the two distributions under H_0 basically overlap with each other. 1000 different watermarks are tested for each distribution. The distributions of the outputs are shown in Fig. 3. A significant improvement is observed by using the LOD.

Then the detection performance is tested under JPEG attacks. The results are shown in Table I. As discussed above, we have m_0=0 and σ_0=1 for both the LCD and the LOD, thus only the (m_1, σ_1) pairs are listed in Table I. JPEG compression under various quality factors is applied, including the no-compression cases, which are the numerical results of those displayed in Fig. 3. From the table we can see, as the quality factor gets lower, the detection performances keep going down, as expected. But LOD always performs better than LCD, in the sense that the (m_1/σ_1) values are always greater by using the LOD.

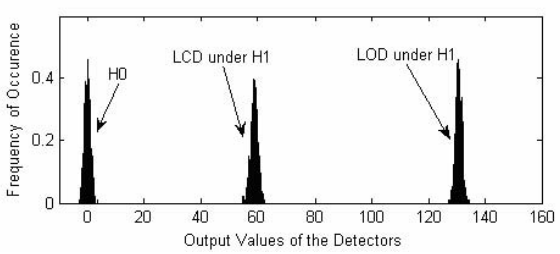
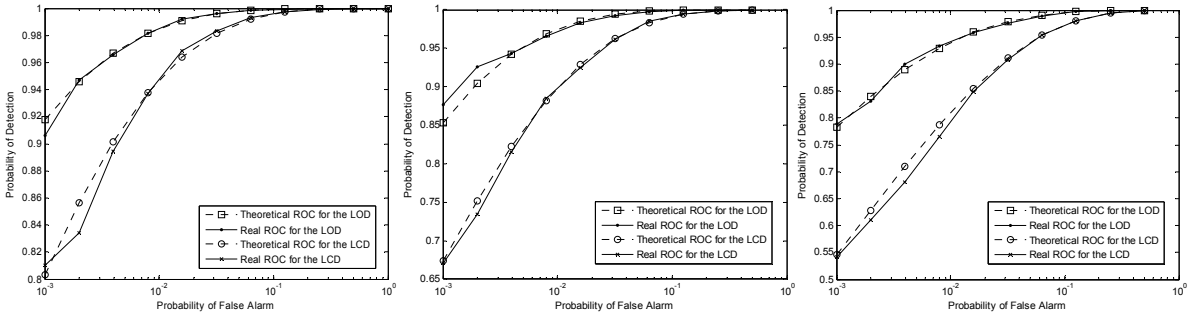


Fig. 3: Output distributions of the detectors

Table I: Comparison of the performances of the LCD and the LOD under JPEG compression. Each Distribution is tested on 1000 watermarks.

Q. Factors		N/A[a]	70	50	30	10
LCD	m_1	58.9	38.1	29.5	22.6	11.3
	σ_1	1.29	1.01	0.92	0.93	0.93
	m_1/σ_1	45.7	37.7	32.1	24.3	12.2
LOD	m_1	131	71.1	50.5	36.9	14.4
	σ_1	1.03	0.89	0.78	0.85	0.90
	m_1/σ_1	127	79.9	64.7	43.4	16.0

[a]N/A means no compression.



(a) (b) (c)

Fig. 4. ROC curves for the 32×32 Lena image after (a) scaling with a factor of 0.3 and adding Gaussian noise with $\sigma_n^2 = 100$; (b) JPEG compression with the quality factor being 30 and (c) JPEG 2000 baseline compression with the bit-rate being 0.5bpp

To plot ROC curves, the Lena image is first sub-sampled to 32×32 before embedding and detection. Otherwise the error rate is too small. Note that in this case the distribution of a detector's output is less Gaussian-like because the CLT needs a large number of samples to take effect. Nevertheless we still use (6) for the theoretical ROC curves, and the comparison with the real ones shows that this is a reasonable approximation. 10000 watermarks are tested for each detector in each case. The results shown in Fig. 4 demonstrate that the LOD always outperforms the LCD.

5. CONCLUSIONS

In this paper, optimum detection is studied for SS watermarks with arbitrary JND models that exploit self-masking. The optimum detector is derived in a closed form, which performs non-linear pre-processing to each observed feature before the correlation with the watermark. We have found the best pre-processing which optimizes the error performance in the sense of Neyman-Pearson criterion. The theoretical analysis is supported by simulation results.

REFERENCES

[1] I. J. Cox, J. Kilian, T. Leighton, and T. Shammoon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. Image Processing*, vol. 6, pp. 1673–1687, Dec. 1997.

[2] S. S. Pradhan, J. Kusuma, and K. Ramchandran, "Distributed compression in a dense microsensor network," *IEEE Signal Processing Magazine*, vol. 19, no. 2, pp. 51–60, Mar. 2002.

[3] C. Podilchuk and W. Zeng, "Image-adaptive watermarking using visual models," invited paper, *IEEE Journal on Selected Areas in Communications,* vol. 16, No. 4, pp. 525–539, May 1998.

[4] J. Hernandez, M. Amado, F.Perez-Gonzalez, "DCT-domain watermarking techniques for still images: detector performance analysis and a new structure," *IEEE Trans. on Image Proc.*, vol. 9, no. 1, Jan. 2000.

[5] A. Nikolaidis and I. Pitas, "Asymptotically Optimal Detection for Additive Watermarking in the DCT and DWT Domains", *IEEE Trans. on Image Processing*, vol. 12, no. 5, pp. 563-571, May 2003.

[6] A. Briassouli and M. G. Strintzis, "Locally Optimum Nonlinearities for DCT Watermark Detection", *IEEE Trans. Image Proc.*, vol. 13, no. 12, Dec. 2004.

[7] Q. Cheng and T. S. Huang, "An additive approach to transform-domain information hiding and optimum detection structure," *IEEE Trans. on Multimedia*, vol. 3, no. 3, pp. 273–284, Sep. 2001.

[8] ——, "Robust optimum detection of transform domain multiplicative watermarks," *IEEE Trans. on Signal Processing*, vol. 51, no. 4, Apr. 2003.

[9] M. Barni, F. Bartolini, A. De Rosa, and A. Piva, "A new decoder for the optimum recovery of nonadditive watermarks," *IEEE Trans. on Image Processing*, vol. 10, no. 5, pp. 755–766, May 2001.

[10] W. Liu, L. Dong, and W. Zeng, "Optimum detection of image-adaptive watermarking in DCT domain," *Intl. Conf. on Image Proc.*, Sept. 2006.

[11] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1994.

[12] W. Liu, L. Dong, and W. Zeng, "Optimum Detection for Spread-Spectrum Watermarking that Employs Self-masking," *(accepted by) IEEE Trans. on Information Forensics and Security*.

[13] C. J. van den Branden Lambrecht, *Perceptual Models and Architectures for Video Coding Applications*, PHD thesis, Swiss Federal Institute of Technology, Aug. 1996.

[14] G. B. Arfken and H. J. Weber, *Mathematical Methods for Physicists*, Orlando, FL, Academic Press, 1985.

[15] K. Sharifi and A. Leon-Garcia, "Estimation of shape parameter for generalized Gaussian distributions in subband decompositions of video", *IEEE Trans. Circuits and Systems for Video Technology*, vol. 5, no. 1, pp. 52–56, Feb. 1995.