A Framework for Optimal Blind Watermark Detection*+

Trista Pei-chun Chen and Tsuhan Chen
Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, U.S.A.
Tel: +1 (412) 268-7536 Email: {peichun;tsuhan}@andrew.cmu.edu

ABSTRACT
We propose a general framework for blind watermark detection. This framework contains a maximum-likelihood detector that utilizes the probability distribution of the original image. Other watermark detectors in literature are shown to be special cases of this framework. We demonstrate this framework in both the pixel domain and the transform domain, and show that our detector outperforms others because of 1) better modeling of the probability distribution of the original image, and 2) consideration to the human visual system in this framework.

Keywords
Blind watermark detection, public watermark, human visual system, maximum-likelihood detection, Gaussian mixture.

1. INTRODUCTION
Watermarking techniques embed ownership information into digital media to protect the copyright [1]-[9]. Watermark techniques can be categorized into two types, *private* and *public* [3]. Private watermarking systems need the original image to detect the watermark, while public watermarking systems can detect the watermark without the original image. In practice, it is better to use public watermarking [8]. Public watermarking is also called blind watermarking [9]. Modern watermarking systems often utilize human visual system (HVS) characteristics [10] to embed the watermark. In this paper, we propose a general framework for blind watermark detection that is based on maximum-likelihood (ML) detection and considers HVS to improve the detection performance. We demonstrate this framework both in the pixel domain and in the discrete cosine transform (DCT) domain. Using this framework, we show that better modeling of the probability distribution function (pdf) of the original image improves the watermark detection performance, compared to conventional Gaussian models [9]. Other modeling effort in literature includes using the Cauchy distribution [12][13] and autoregressive [4] models.

* Work supported in part by Institute for Information Industry

+ Published in Proc. of ACM Multimedia 2001 Workshop on Multimedia and Security: New Challenges, Ottawa, Canada, October 2001

2. BLIND WATERMARK DETECTION
2.1 General Framework
A watermark is typically represented by a number of bits. To embed the watermark into the original image, we alter some properties, such as pixel values or transform coefficients, of the original image to embed each bit of the watermark. To improve robustness, the same watermark bit can be embedded multiple times such as:

$$\begin{aligned} H_1 &: y_i = x_i + w_i \\ H_0 &: y_i = x_i - w_i \end{aligned} \quad i = 1, \ldots, N \tag{1}$$

where y_i represents the received watermarked signal, x_i represents the original signal, and w_i represents the watermark amplitude. Watermark detection is hence a hypothesis-testing problem between two hypotheses H_1 and H_0: H_1 if 1 is embedded, and H_0 if 0 is embedded. Note that for blind watermarking, the original signals comprise the major *noise* that affects the watermark detection performance. Other types of noise usually have less power and hence less impact.

To take HVS characteristics into consideration, we let the watermark amplitude to be a general function, $f_i(x_i)$, of the original signal x_i. This is an extension to [5] that assumes this to be a linear transformation of the original signal. In addition to the dependency on x_i, in general $f_i(\cdot)$ can be different for each of the N signals, so

$$\begin{aligned} H_1 &: y_i = x_i + f_i(x_i) \equiv g_{i,1}(x_i) \\ H_0 &: y_i = x_i - f_i(x_i) \equiv g_{i,0}(x_i) \end{aligned} \quad i = 1, \ldots, N \tag{2}$$

We represent the watermarked signal with 1 embedded as $y_i = g_{i,1}(x_i)$ and the watermarked signal with 0 embedded as $y_i = g_{i,0}(x_i)$. Given the received watermarked signals, maximum *a posterior* (MAP) detection can be performed to determine the embedded watermark bit. The MAP detection is simplified to maximum-likelihood (ML) detection if the *a priori* probabilities for both hypotheses are equal, i.e., $p(H_1) = p(H_0)$. Assuming that the selection of the N signals are random enough so that y_i's are independent with each other conditioned on H_1 (and H_0), ML detection can be written as:

$$\prod_{i=1}^{N} \frac{p(y_i|H_1)}{p(y_i|H_0)} \begin{array}{c} H_1 \\ \geq \\ < \\ H_0 \end{array} 1 \qquad (3)$$

With some manipulation of the above, we can derive the optimal watermark detector as follows:

$$\left[\sum_{i=1}^{N} \{\ln[p(g_{i,1}^{-1}(y_i))] - \ln[p(g_{i,0}^{-1}(y_i))]\} + \sum_{i=1}^{N} \left(\ln|J_{i,1}(y_i)| - \ln|J_{i,0}(y_i)|\right) \right] \begin{array}{c} H_1 \\ \geq \\ < \\ H_0 \end{array} 0 \qquad (4)$$

where $J_{i,1}(y_i) \equiv \left| \frac{d[g_{i,1}^{-1}(y_i)]}{dy_i} \right|$ and $J_{i,0}(y_i) \equiv \left| \frac{d[g_{i,0}^{-1}(y_i)]}{dy_i} \right|$.

2.2 Optimal Watermark Detector

The proposed optimal watermark detector in (4) has two features: 1) it uses better modeling of the pdf of the original image, and 2) it incorporates the HVS into the detection process. Note that with blind watermark detection, we do not have information about the original image. However, knowledge about the *statistics* of the original image, such as the pdf of the pixel values or the pdf of the transform coefficients, allows us to detect the watermark with better performance. In practice, the detector would need to estimate the pdf of the original image from the received watermarked image. Such estimation is usually good, since the watermarked image is similar to the original image as the watermark is typically chosen to be imperceptible. Note that terms $J_{i,1}(\cdot)$ and $J_{i,0}(\cdot)$ in (4) incorporate HVS into the framework. Conventional watermark detectors assume the watermark being independent of the original image, in which case these two terms are both equal to unity.

3. PIXEL-DOMAIN WATERMARKING

3.1 Gaussian Mixture Modeling

We first demonstrate the proposed framework using pixel-domain watermarking. In this case, we need to know the pdf of the pixel values of the original image. As an example, Figure 1 shows the histogram of pixel values in the image Lena. We see that such a distribution cannot be modeled as a single Gaussian function. Therefore, conventional watermark detectors that assume a Gaussian pdf usually do not perform well [9].

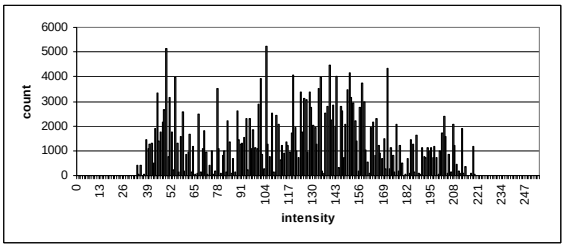


Figure 1. Histogram of the pixel values of Lena.

We propose to model the pdf as a Gaussian mixture with M components:

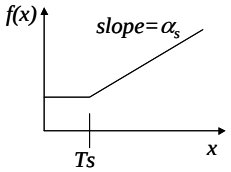
$$p(x) = \sum_{j=1}^{M} w_j N_{(m_j, \sigma_j^2)}(x) \qquad (5)$$

and the optimal watermark detector can be obtained by substituting (5) into (4).

Note that the number of components M affects the modeling performance significantly. If M is too small, the Gaussian mixture cannot model the histogram well. If M is too big, the Gaussian mixture may over-fit the histogram and stray away from the desired pdf. We determine M as follows. Starting from one component, i.e., $M = 1$, we increment M by one at a time, and estimate the pdf. We test the correlation (or the relative entropy [14]) between the newly estimated pdf and the histogram. If it is greater than the correlation between the old estimated pdf and the histogram by more than a threshold ε_ρ, we continue increasing M. Otherwise, we stop and use the current value of M.

3.2 HVS in the Pixel Domain

In the pixel domain, we design the watermark amplitude $f(x_i)$ based on the Weber's law as follows:



$$f(x) = \alpha_S \max\{x, T_S\} \qquad (6)$$

The two parameters α_S and T_S are adjustable depending on the imperceptibility requirement. We use $T_S = 20$ and $\alpha_S = 0.05$ in our experiments and verify that the resulting watermarked image has no visible artifacts.

3.3 Experiments

To show how different pdf modeling methods affect the watermark detection performance, four methods are compared: (a) single Gaussian pdf; (b) Gaussian mixture with two components; (c) Gaussian mixture with the number of components determined as described in Section 3.1 with $\varepsilon_\rho = 0$; (d) Gaussian mixture with the number of components determined as described in Section 3.1 with $\varepsilon_\rho = 0.005$.

The three test original images are 8-bit gray-level images of size 512×512 as shown in Figure 2. The 10 test watermarks are 4-bit gray-level images of size 8×8, each carrying 256 bits of information. All watermarks are chosen so that the *a priori* probabilities for both hypotheses are equal. With all combinations of originals and watermarks, there are 30 sets of data. The experiment results shown in the following are the average of these 30 sets of data.

Figure 2. Test original images.

Each bit of the watermark is distributed randomly according to a key to $N = 1024$ pixels of the original image. The amount by which the original pixel is altered is determined by (6), and the resulting watermarked images have an average PSNR of 32.0dB. At the watermark detector, each bit of the watermark is detected by performing ML detection as shown in (4). For the four methods, (a)-(d), the bit error rate (BER) of the retrieved watermark at various quality factors of JPEG compression [15] is shown in Figure 3.

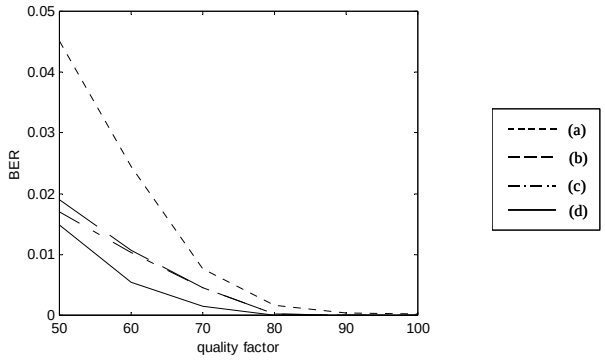


Figure 3. BER of pixel-domain watermark detectors.

We can see that the proposed watermark detector (d) with Gaussian mixture modeling and $\varepsilon_\rho = 0.005$ gives the best result. The conventional watermark detector (a), which assumes one single Gaussian, gives the worst performance. Watermark detector (b), which assumes two Gaussian components, gives slightly better performance than (a). Watermark detector (c), with $\varepsilon_\rho = 0$, over-fits the histogram with too many components and gives worse performance than the proposed watermark detector (d).

4. DCT-DOMAIN WATERMARKING

4.1 Generalize Gaussian Modeling

If the watermark is embedded and detected in the DCT domain, we need to know the pdf of the DCT coefficients of the original image. As an example, Figure 4 shows the histogram of the fifth ac coefficients of the image Lena.

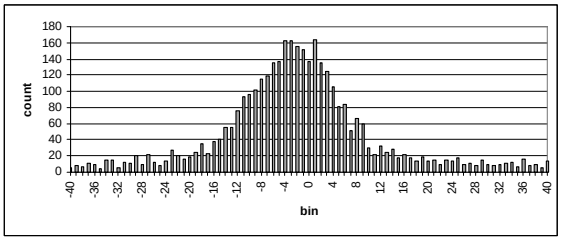


Figure 4. Histogram of the fifth ac coefficients of Lena.

Such a distribution can be modeled as a *generalized Gaussian distribution* [16]:

$$p(x) = Ae^{-|\beta x|^c};\ \beta = \frac{1}{\sigma}\sqrt{\frac{\Gamma(3/c)}{\Gamma(1/c)}};\ A = \frac{\beta c}{2\Gamma(1/c)} \tag{7}$$

where σ is the standard deviation of the DCT coefficient x, c is the generalized Gaussian distribution parameter, and $\Gamma(\cdot)$ is the Gamma function. The optimal watermark detector can be obtained by substituting (7) into (4). This is more general than the watermark detector proposed by Hernández *et al*'s [16]. In particular, if we let $f_i(x_i)$ be a constant value d_i, $g_{i,1}(\cdot)$ and $g_{i,0}(\cdot)$ become linear functions and the derivatives of $g_{i,1}^{-1}(\cdot)$ and $g_{i,0}^{-1}(\cdot)$ become unity, i.e., $J_{i,1}(\cdot) = J_{i,0}(\cdot) = 1$. In this case, our watermark detector degenerates into Hernández's watermark detector:

$$\sum_{i=1}^{N}\left(\left|\beta_i(y_i + d_i)\right|^{c_i} - \left|\beta_i(y_i - d_i)\right|^{c_i}\right) \underset{H_0}{\overset{H_1}{\gtrless}} 0 \tag{8}$$

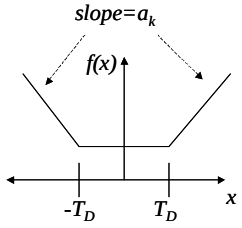
Further, if the pdf of the original DCT coefficients is modeled as a simple Gaussian distribution, (8) becomes:

$$\sum_{i=1}^{N} y_i d_i \underset{H_0}{\overset{H_1}{\gtrless}} 0 \tag{9}$$

which is the correlation detection algorithm used by Zeng and Liu [8]. Therefore, we see that both watermark detectors of Hernández's and Zeng's are special cases of the proposed watermark detection framework.

4.2 HVS in the DCT Domain

In the DCT domain, there are two major characteristics of HVS, *frequency selectivity* and *contrast sensitivity*. To utilize these, we choose the watermark amplitude as follows:



$$f_k(x) = a_k \cdot \max(T_D, |x|) = e_k \alpha_D \cdot \max(T_D, |x|) \tag{10}$$

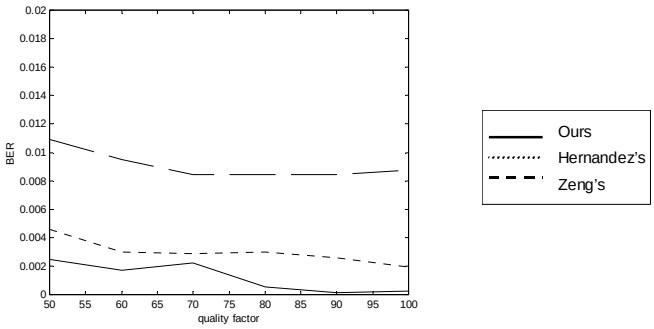
where e_k are determined by the frequency selectivity. We use $T_D = 10$ and $\alpha_D = 0.025$ in our experiments and verify that the resulting watermarked image has no visible artifacts.

4.3 Experiments

We use (10) to generate the watermarked images to compare our approach with Hernández's and Zeng's approaches. The watermarked images have an average PSNR of 32.2dB. We then

detect the watermark using the three approaches. In order to calculate the watermark amplitude and to get the statistics of the original DCT coefficients, all three detectors use the received watermarked DCT coefficient y as an estimate of the original DCT coefficient x. In Hernández's and Zeng's approaches, the d_i used is also estimated. The watermark embedding and detection are performed in 8×8 blocks of the original and watermarked image respectively. The test images are the same as those in Section 3.

The resulting BER of the watermark retrieved by the three detectors, at different quality factors of JPEG, is shown in Figure 5.



(a) (b)

Figure 5. BER of DCT-domain watermark detectors.

We see that our proposed DCT domain watermark detector performs the best among the three detectors because it has the best modeling of the pdf of the DCT coefficients. Comparing Hernández's with Zeng's approaches, the former is slightly better since it uses generalized Gaussian distribution that is better than using a simple Gaussian distribution.

5. Conclusion

In this paper, we provided a general framework for blind watermark detection and proposed an optimal watermark detector based on ML detection and modeling of the pdf of the original image. We showed that other watermark detectors are special cases of this general framework. We also showed that better modeling of the pdf of the original image resulted in better watermark detection performance, both in the pixel domain and the DCT domain. In pixel-domain watermarking, we modeled the pdf of the pixel values of the original image as a Gaussian mixture with a well-chosen number of components. In DCT-domain watermarking, we modeled the pdf of the DCT coefficients of the original image as a generalized Gaussian distribution.

Our future work includes better modeling of the original image in other domains, such as the discrete wavelet transform domain. Further, we plan to incorporate noise terms other than the original image, e.g., the quantization error, similar to the work done in [4]. With more sophisticated modeling, watermark detection can be robust to various types of noises.

6. REFERENCES

[1] G. C. Langelaar, I. Setyawan, and R. L. Lagendijk, "Watermarking Digital Image and Video Data". *IEEE Signal Processing Magazine*, September 2000, 20-46.

[2] F. Hartung, and M. Kutter, "Multimedia Watermarking Techniques". *Proceedings of IEEE*, 87(7), 1079-1107.

[3] R. B. Wolfgang, C. I. Podilchuk, and E. J. Delp. "Perceptual Watermarks for Digital Images and Video". *Proceedings of IEEE*, 87(7), 1108-1126.

[4] I. J. Cox, M. L. Miller, and A. L. McKellips, "Watermarking as Communications with Side Information". *Proceedings of IEEE*, 87(7), 1127-1141.

[5] J. R. Hernández and F. Pérez-González, "Statistical Analysis of Watermarking Schemes for Copyright Protection of Images". *Proceedings of IEEE*, 87(7), 1142-1166.

[6] G. Voyatzis and I. Pitas, "The Use of Watermarks in the Protection of Digital Multimedia Products". *Proceedings pf IEEE*, 87(7), 1197-1207.

[7] M. D. Swanson, M. Kobayashi, and A. H. Tewfik, "Multimedia Data-Embedding and Watermarking Techniques". *Proceedings of IEEE*, 86(6), 1064-1087.

[8] W. Zeng and B. Liu, "A Statistical Watermark Detection Technique Without Using Original Images for Resolving Rightful Ownerships of Digital Images". *IEEE Transactions on Image Processing*, 8(11), 1534-1548.

[9] J. J. Eggers, J. K. Su; B. Girod, "Robustness of a blind image watermarking scheme". *Proceedings of ICIP2000*, 17-20.

[10] Andrew B. Watson, *Digital Images and Human Vision*, Massachusetts Institute of Technology 1993.

[11] J. R. Hernández, F. Pérez-González, J. M. Rodriguez, and G. Nieto, "Performance analysis of a 2-D-multipulse amplitude modulation scheme for data hiding and watermarking of still images". *IEEE Journal on Selected Areas in Communications*, 16(4), 510-524.

[12] E. Sayrol, J. Vidal, S. Cabanillas, and S. Santamaria, "Optimum watermark detection in color images". *Proceedings of ICIP1999*, 231-235.

[13] J. Vidal and E. Sayrol, "Optimum watermark detection and embedding in digital images". *Proceedings of ICIP1999*, 285-290.

[14] Thomas M. Cover, and Joy A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Inc. 1991.

[15] Independent JPEG Group website—http://www.ijg.org/.

[16] J. R. Hernández, M. Amado, and F. Pérez-González, "DCT-Domain Watermarking Techniques for Still Images: Detector Performance Analysis and a New Structure". *IEEE Transactions on Image Processing*, 9(1), 55-68.