Robust Optimum Detection of Transform Domain Multiplicative Watermarks

Qiang Cheng, *Member, IEEE*, and Thomas S. Huang, *Life Fellow, IEEE*

***Abstract*—Digital watermarking is an emerging technique to protect data security and intellectual property right. Identification or verification of watermarking patterns can be achieved by detecting watermarks in received signals. However, one of the biggest challenges in watermarking detection is that the strengths of the watermark signals will change after being distorted by an attacker in a watermarking channel. Meanwhile, the embedding strengths may be adapted to original signals, which are unknown at the receiver end. Further, the original signals are often highly non-Gaussian. Although some work has been done on optimum detection of watermarks, the uncertainty of watermark signal strengths and real statistical behavior of multimedia contents have not been taken into account simultaneously. Much more study is needed to enhance the performance of watermarking systems. Since multiplicative watermarks are robust and well suited for copyright protection, this paper presents our investigation on robust optimum detection of multiplicative watermarks. For sub-band transformed domains such as the discrete cosine transform (DCT), discrete wavelet transform (DWT), and pyramid transform, a class of generalized correlators is constructed based on the generalized Gaussian distributions. Thresholding methods to achieve a given false alarm rate, and the performance analyses are provided. The square-root detector is designed and demonstrated to have near optimal performance for a large set of natural images and can be employed as a "universally optimal" detector or decoder for images and video. The locally most powerful detection method is then extended to DFT domain multiplicative watermarking, with the magnitudes of coefficients modeled by the Weibull distributions. Another class of detectors is built based on this statistical modeling. The robust optimum detection of multiplicative watermarks can be applied to copyright notification, enforcement, and broadcast monitoring. We have applied the robust optimum watermarking detection to combined audio and video watermarking.**

***Index Terms*—Digital watermarking, generalized correlation detector, generalized Gaussian distribution, robust optimum detection, square-root detector, Weibull distribution.**

I. Introduction

MULTIMEDIA watermarking has attracted increasing interest from many areas as the data security and copyright protection issues are becoming increasingly important [1], [2]. It embeds hidden secondary data into digital multimedia products for copyright notification and protection, content authentication, transaction-tracing, and covert communication. The main advantage of watermarking is that it provides a way to deliver side information through primary multimedia contents in a seemingly innocuous and standards-compliant fashion, such that many novel functionalities can be enabled. In steganography or security applications, secret messages may be transmitted covertly through a perceptually innocent image or audio. In multimedia database retrieval, watermarking patterns associated with annotations or keywords may be imprinted seamlessly into host media to facilitate future accurate access. In broadcast monitoring and copy control techniques, watermarking can actively and cost-effectively identify specific multimedia contents in digital TV, audio, or video broadcasting or playing back such that royalty collection can be automated or illegal copying prevented.

Since the inception of digital watermarking around the early 1990s, there have been a variety of methods proposed in the literature, and there are many ways to classify them. For example, some approaches deal with the signals in the sample (spatial or time) domain, while others deal with transformed data. Some private watermarking schemes need the knowledge of host multimedia signals in decoding, whereas blind watermarking schemes do not. We will view watermarking as the following information system with side information available only to the embedder [3]–[7]. A secret message or pattern is encoded by an encoder or embedder into a watermark and hidden into the host medium within an embedding distortion level. The composite signal is then input into the watermarking channel, where an attacker attempts to disrupt the watermark by introducing additional distortions. The channel output is a corrupted or noised composite signal. A decoder decodes the watermark bit by bit or symbol by symbol, or a detector detects or verifies the existence of a specific watermarking pattern or specific message.

According to the roles of watermarks to play in this information system, the existing schemes are classified into two categories. In the first category, watermarks serve as transmission codes, where an ordinary communication channel model is used, and where a message is encoded by the embedder and the decoder carries out full decoding to extract the message [3]–[6]. The embedding often needs to partition or enforce the host signal space into subsets that are mapped to the values taken by the watermark. For example, sign enforcement [8], [9] and table-lookup [10], [11] are two basic ways to do the partition. More sophisticated approaches incorporate the Costa scheme [12] for high channel capacity, and quantization is exploited for set partitioning [6], [13]. These schemes have

Manuscript received February 4, 2002; revised December 3, 2002. This work was supported in part by the National Science Foundation under Grants CDA 96-24386 and 96-24396. The associate editor coordinating the review of this paper and approving it for publication was Dr. Bede Liu.

Q. Cheng is with the Electrical and Computer Engineering Department, Wayne State University, Detroit, MI 48202 USA (e-mail: qcheng@ece.eng.wayne.edu).

T. S. Huang is with the Electrical and Computer Engineering Department and Beckman Institute, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: huang@ifp.uiuc.edu).

Digital Object Identifier 10.1109/TSP.2003.809374

high rates, and the original host signals will not form a noise source to interfere with the decoding. However, these schemes are usually not very robust [13] since the robustness of these schemes depends on tolerance zones or minimum intersymbol distances that are always very limited due to the perceptual quality requirement [13].

In the second category, watermarks serve as verification codes, where full decoding is not really necessary, but only to decide whether or not *a particular* message or hypothesized pattern is present or has been transmitted. Thus, a binary decision is often required. In copyright protection or notification and broadcast monitoring, such verification watermarking usually suffices [7]. Remarkably, a double exponential number of messages can be verified reliably in a verification channel. Instead, an exponential number of messages can be reliably transmitted in an ordinary watermarking transmission model.

In most watermarking verification schemes, without deterministic mapping from set partition of host signal space to watermark symbols, host signals are often modeled as noise, and many approaches have been taken to suppress noise from both host signals and attacks to improve detection effectiveness. For example, the additive spread spectrum scheme spreads the watermark signal to many components [14], and the maximum watermark strengths are obtained from just-noticeable-differences (JND) [15]–[17]. Since a number of samples collectively contribute to the detection of a watermark pattern, such verification watermarking usually has high robustness, even though its ability to recover more than one information bit might not be as great as that of transmission watermarking. Thus, it is more suitable for applications such as copyright protection or notification, copy control, device control, and broadcast monitoring.

Identification or verification of watermarking patterns can be achieved by detecting the watermarks in the received signals [7], [18]. However, one of the biggest challenges in watermarking detection is that the strengths of the watermark signals would change after being distorted by the attacker in the watermarking channel. Meantime, the embedding strengths may be adapted to the original signals that are unknown at the receiver end. Further, the original signals are often highly non-Gaussian. Although some work has been done on optimum detection of watermarks, robust optimum detection still needs much study. In the literature, the watermarks are often embedded in either an additive way or a multiplicative way. An optimum detection structure for additive watermarks has been derived [20]–[23], and accurate performance analyses have been developed based on the generalized Gaussian distributions (GGDs). Multiplicative watermarks are automatically image content dependent [19] and are automatically embedded mainly into the perceptually most significant components of the image. The perceptual models based on Weber's law can be easily exploited. Due to these desired properties, multiplicative watermarks have considerable robustness, and they are well-suited for copyright protection. For multiplicative watermarks, a correlation detector is often used [14], [19]. It is simple and intuitively appealing; however, its employment for the detection of multiplicative watermarks cannot be justified, and in the discrete Fourier transform (DFT) domain, a new detection structure has been derived [24], [25]. In Section II, it is to be shown that the correlator actually cannot work for multiplicative watermarks in many sub-band transformed domains such as the discrete cosine transform (DCT), discrete wavelet transform (DWT), and pyramid transforms. For multiplicative watermarks embedded in the magnitudes of DFT coefficients, the correlator can work, but it is optimal only when the magnitudes follow the exponential distribution.

To obtain the best performance, optimum detection of the multiplicative watermarks needs to be investigated. In this paper, a class of robust optimum detection statistics taking into account perceptual masking effects and attacks is derived, based on the generalized Gaussian distributions for multiplicative watermarks in sub-band transformed domains such as DCT, DWT, and pyramid transforms. The performance of the proposed detector is examined using the ordinary Gaussian distribution. The commonly used Laplacian distribution is a special case of the GGD law, and the optimum detector corresponding to it is also examined in this paper. Their performances are compared with the optimum detector corresponding to shape parameter $1/2$, which is identified as a "universally optimum" detector for its near-optimal performance for a large set of natural images. The methodology is then extended to multiplicative watermarks embedded in the magnitudes of DFT coefficients. Another class of optimum detectors is obtained by using the Weibull distribution. The threshold achieving a given false alarm probability is provided. The effectiveness of the optimum detection is demonstrated by extensive experiments. The robust optimum detection for multiplicative watermarks can be applied to copyright notification, enforcement, and broadcast monitoring. In this paper, an application of the optimum detector to combined audio and video watermarking is proposed for the purpose of content authentication.

The rest of this paper is organized as follows. Multiplicative embedding rule is introduced in Section II. The robust optimum detection rule for GGD is derived in Section III. The methodology is extended to DFT-domain multiplicative watermarking in Section IV. Experiments are conducted in Section V to validate the analysis. Section VI applies the optimum detector to the combined audio and video watermarking. Section VII concludes the paper.

II. Multiplicative Watermark Embedding Rule and Detections

A. Multiplicative Embedding Rule and Correlator

For identification or verification watermarking channel, a watermark pattern is embedded at the transmission end. In the literature, the embedding makes use of either an additive watermark embedding rule or a multiplicative one.

The commonly used additive embedding rule is [14], [19]

$$y_i = x_i + \gamma w_i, \qquad i = 1, \ldots, N \tag{1}$$

where $\mathbf{x} = \{x_1, \ldots, x_N\}$ is a sequence of data from the (transformed) original image, $\mathbf{w} = \{w_1, \ldots, w_N\}$ is a sequence of watermark signals, γ is a gain factor, and $\mathbf{y} = \{y_1, \ldots, y_N\}$ is a sequence of watermarked data. Using the same notation, the commonly used multiplicative embedding rule is [14], [19]

$$y_i = x_i(1 + \gamma w_i) \qquad i = 1, \ldots, N. \tag{2}$$

For detection or verification, the receiver needs to verify if *a specific* watermarking pattern exists or not. For full decoding of a message similar to ordinary communication, orthogonal modulation combined with code division multiple access (CDMA) can be applied [26], [27]. To these ends, a correlator is often used [19]:

$$R_{\mathbf{y}\cdot\mathbf{w}} = \frac{1}{N}\sum_{i=1}^{N} y_i w_i \geq T \quad \Rightarrow \quad \mathbf{w} \text{ is present}$$
$$< T \quad \Rightarrow \quad \mathbf{w} \text{ is not present} \qquad (3)$$

where T is a threshold. For additive watermarks, the correlator can work, but it is optimal only if the distribution is Gaussian. Optimum detectors as well as the performance analysis for additive watermarks have been investigated [18], [20]–[23]. For multiplicative watermarks embedded in many sub-band transformed domains, such as DCT, DWT, and pyramid transforms, it is shown as follows that this correlator actually cannot work at all when the watermark is independent of the original signal. Indeed, for sub-band coefficients x_i, $i = 1, \ldots, N$, consider the verification of $\{w_i\}$. When there is no channel attack at all, $y_i = x_i$ if there is no watermark; then, $R_{\mathbf{y}\cdot\mathbf{w}} = (1/N)\sum_{i=1}^{N} x_i w_i \approx 0$ since x_i is zero mean; on the other hand, $y_i = x_i + \gamma x_i w_i$ if there is such a watermark; then, $R_{\mathbf{y}\cdot\mathbf{w}} = (1/N)\sum_{i=1}^{N} x_i w_i + (1/N)\sum_{i=1}^{N} \gamma x_i w_i^2 \approx 0$. A simple fix to this problem is to change the embedding rule but keep the detection rule. For example, the following rules can be used: $y_i = x_i(1 + \gamma\, sign(x_i) w_i)$ or $y_i = |x_i|(1 + \gamma w_i)$, where $sign(x_i)$ is the signum function of x_i, namely, it is 1 when $x_i \geq 0$; otherwise, it is -1. In this way, the correlator can work but it is still not optimum. Another fix is to keep the embedding rule but change the detection rule. In doing so, we are able to investigate the optimum detection for multiplicative watermarks embedded in many sub-band transformed domains. The methodology is also extended to multiplicative watermarks embedded in the magnitudes of DFT coefficients. It is demonstrated that the correlator is outperformed by the optimum detectors derived in this paper.

B. List of New Optimal Detection Structures

For the convenience of making comparisons, new optimum detectors constructed in this paper are listed in the following. The details of the development of these detection structures are in Sections III and IV.

For multiplicative watermarks embedded into the sub-band transformed domains, such as DCT, DWT, and pyramid transforms, if the embedding depths γ_i^* are known to the detector, the uniformly most powerful detector is

$$L(\mathbf{y}) = \sum_{i=1}^{N} |\beta_i y_i|^{c_i} \left(1 - (1 + \gamma_i^* w_i)^{-c_i}\right)$$

where β_i and c_i are the parameters of GGD (see Section III-A).

Taking account of visual modeling and attacks, the robust optimum detector is

$$\delta(\mathbf{z}) = \frac{1}{N}\sum_{i=1}^{N} c_i |\beta_i z_i|^{c_i} w_i$$

where $\mathbf{z} = \{z_i\}_{i=1}^{N}$ is a sequence of output of the watermarking channel (see Section III-C).

To achieve near-optimal performance for a large set of natural images or video frames, a "universally optimal" detector is

$$\delta(\mathbf{z}) = \frac{1}{N}\sum_{i=1}^{N} |z_i|^{1/2} w_i$$

which is particularly simple in form and is a special case of the robust optimum detector (see Section III-C).

For multiplicative watermarks embedded into the DFT magnitudes, the robust optimum detector is

$$\delta(\mathbf{z}) = \sum_{i=1}^{N} \rho_i \left(\frac{z_i}{\alpha_i}\right)^{\rho_i} w_i$$

where α_i and ρ_i are the parameters of the Weibull distribution (see Section IV).

Derivations and appropriate thresholds for these detection structures are developed in the following sections. Computational complexities of these detectors are compared with that of the correlator in Section III-C. The performance of the robust optimal detector is compared with that of the uniformly optimal detector in 5, where the performance evaluations of the robust optimal detectors are also conducted.

III. Optimum Detection for Multiplicative Watermarks

For additive watermarks, the optimum detector has been derived, and the performance analysis has been conducted based on the GGD model [20]–[23]. To obtain the best performance for multiplicative watermarks, we analyze the optimum detection of the multiplicative watermarks and provide novel optimum detectors effective to various transformed domains. Specifically, the optimum detection statistic is derived for a class of general distributions, namely, the generalized Gaussian distributions. The methodology is extended to multiplicative watermarks cast in the magnitudes of DFT coefficients, where the magnitudes may be statistically described using the Weibull distribution.

A. GGD Model for Sub-band Coefficients

Multiplicative watermarks are often embedded in the transformed domains, such as DFT, DCT, DWT, and pyramid transforms. The coefficients of DCT, DWT, and Pyramid transform can be statistically modeled using GGD. Optimum detection for additive watermarks using GGD has been studied [20]–[22], and large deviation bounds for performance analysis have also been developed [22], [23]. Many watermarking techniques in the DFT domain employ amplitude modulation, taking advantage of the translation or shift invariance. GGD is not suitable for these techniques since DFT magnitudes are positive. We proceed by first modeling sub-band transform coefficients using a global stationary GGD and construct both uniformly most powerful detectors and robust optimum detectors; then, we model the magnitudes of DFT coefficients using Weibull distribution and constructing corresponding robust optimum detectors.

The probability density function (pdf) of the generalized Gaussian distribution is

$$p_X(x) = Ae^{-|\beta(x-m)|^c} \tag{4}$$

where

$$\beta = \frac{1}{\sigma}\left(\frac{\Gamma(3/c)}{\Gamma(1/c)}\right)^{1/2}, \qquad A = \frac{\beta c}{2\Gamma(1/c)}$$

and m is the mean, σ is the standard deviation of the distribution, and $\Gamma(t) = \int_0^\infty u^{t-1}e^{-u}\,du$ is the Gamma function. The distribution is denoted as $GGD(c;\, m,\, \sigma)$. Many image transforms are orthogonal transforms and the mean value of coefficients in the AC sub-bands is close to zero, i.e., $m = 0$ [28]. The power of the exponent c is the shape parameter. The smaller c is, the more impulsive the shape, and the heavier the tails. The GGD model contains the Laplacian and the Gaussian distributions as special cases, with $c = 1$ and $c = 2$, respectively. When $c \to 0$, it approaches a Dirac impulse. When $c \to \infty$, it tends to the uniform distribution. It has been found out that for AC coefficients in low or middle frequency sub-bands, generally $0.4 < c \leq 1$, with most images having c around 0.5. The imperceptibility requirement of watermarking implies that the magnitudes of watermark signals have to be small compared with the original image, and the total watermark energy should not be very large either. In this typical weak-signal scenario, good estimations of c and σ can be obtained from the watermarked image instead of the original. For example, the values of c and σ for the original Lena are 0.48 and 11.80; for watermarked Lena using multiplicative embedding rule with $\gamma = 0.20$ and after JPEG compression with quality factor 75, the estimated values are $c = 0.47$ and $\sigma = 11.78$. The differences from those of the original image are indeed insignificant. In cases where the estimation of parameters is undesirable, a "universally optimal" detector identified in Section III-C can be utilized. It does not need to estimate the parameters for each image yet achieves near optimal performance for a large set of natural images.

A globally stationary GGD can be employed to model the sub-band coefficients, where the distribution has a single variance. An alternative is to make use of a nonstationary Gaussian model, where a model of parallel Gaussian channels is used, with zero mean and with variance of each Gaussian channel estimated individually. In the parallel Gaussian model, the host signal is segmented into subsignals, and each subsignal whose size is relatively small is modeled approximately by a stationary Gaussian model. Since a certain subsignal is always difficult to model using the Gaussian distribution even if the size becomes small—for example, a subimage containing an edge on a simple background will typically have a high peak in the histogram which the Gaussian model does not have [29]—stationary GGD can be used to provide a more reasonable model for sub-band coefficients even for signals with small size.

B. Uniformly Most Powerful Detector

The gain factor γ for the multiplicative rule in (2) is small due to perceptual constraints. The multiplicative embedding rule relies on Weber's law as its perceptual model [19], according to which the gain factor of noise can be as high as 2% of the local luminance without being perceived by human eyes [28]. Because of other masking effects such as the frequency sensitivity and the contrast masking, the actual value of γ may be even higher than 0.02 without incurring perceptual degradation. For stronger robustness or higher data rates, bigger γ can be applied. However, if γ is too big, the perceptual quality will degrade. The embedding process may depend on multimedia content and employ visual masking models to vary γ. In adddition, different γs may be adopted, for example, from frame to frame in video watermarking. It is reasonable to use

$$y_i = x_i(1 + \gamma_i w_i) \tag{5}$$

where γ_is are different embedding strengths for different x_is. Assuming the actual embedding strengths γ_i^* are known to the detector, a simple hypothesis testing is as follows:

$$\tilde{H}_0\colon \gamma_i = 0 \quad \text{versus} \quad \tilde{H}_1\colon \gamma_i = \gamma_i^*. \tag{6}$$

Denote the pdf under $\tilde{H}_\theta$ as p_θ, $\theta = 0, 1$. For GGD laws, $x_i \sim \text{GGD}(c_i;\, 0,\, \sigma_i)$. The likelihood ratio test (LRT) leads to

$$l(\mathbf{y}) = \frac{p_1(\mathbf{y})}{p_0(\mathbf{y})} = \prod_{i=1}^{N}(1 + \gamma_i^* w_i)^{-1} \cdot \exp\left\{\sum_{i=1}^{N} |\beta_i y_i|^{c_i}\, \frac{(1+\gamma_i^* w_i)^{c_i} - 1}{(1+\gamma_i^* w_i)^{c_i}}\right\}. \tag{7}$$

By taking logarithm, it can be simplified into

$$L(\mathbf{y}) = \sum_{i=1}^{N} |\beta_i y_i|^{c_i}\left(1 - (1+\gamma_i^* w_i)^{-c_i}\right) \tag{8}$$

and the optimum test in the sense of the Neyman–Pearson lemma is

$$\begin{aligned} L(\mathbf{y}) > T^* \quad &\Rightarrow \quad \tilde{H}_1 \\ < T^* \quad &\Rightarrow \quad \tilde{H}_0 \end{aligned} \tag{9}$$

where T^* is a properly chosen threshold to maximize the detection probability for a given false alarm rate. Its performance can also be investigated in a way similar to additive watermarking in [22] and [23]. However, other optimum detection structures are pursued, which are optimal in the sense of locally most powerful detection [30].

C. Robust Optimum Detection of Multiplicative Watermarks

Attacks usually change the gain values. In the watermarking channel, an attacker attempts to disrupt the watermark while preserving the image quality. The attacker introduces a source of noise, and the output of the attacking channel $\mathbf{z} = \{z_k\}_{k=1}^N$ becomes

$$z_k = y_k + e_k = x_k + \gamma_k w_k x_k + e_k, \qquad k = 1, \ldots, N \tag{10}$$

where $\{e_k\}_{k=1}^N$ is the attack noise. Decomposing the noise as $e_k = w_k x_i \tilde{\gamma}_k$, the watermark strengths after attacks are

$$\gamma_k' = \gamma_k + \tilde{\gamma}_k, \qquad k = 1, \ldots, N. \tag{11}$$

If $\tilde{\gamma}_k$ is positive, the attacker helps to increase the embedding strength. In the case of robust watermarking, the largest gain factor γ_k could be determined using the just noticeable difference (JND), representing the largest permissible modulation strength without incurring perceptual degradation. For positive $\tilde{\gamma}_k$, γ'_k exceeds JND, and the perceptual quality of the image is degraded. An ideal attacker, which is the worst case for the embedder, may introduce negative $\tilde{\gamma}_k$ to decrease the embedding strengths or better obliterate it. To account for the attacks, we set a tolerance zone and seek the robust optimum detection of multiplicative watermarks.

Assimilating attacks as well as visual masking effects, the following composite hypothesis testing is considered to derive robust optimum detectors:

$$H_0\text{: } \gamma'_k = 0 \quad \text{v.s.} \quad H_1\text{: } \gamma'_k > 0, \qquad k = 1, \ldots, N. \tag{12}$$

If the attacking channel is assumed known to the embedder in some sense, a watermarking detection game may be formulated, where the attacker attempts to maximize the probability of detection errors, whereas the embedder minimizes it, subject to certain constraints. Often, it is difficult to find a closed-form solution to this minimax optimization problem except for some attacking channels with certain strict restrictions [31]. For general attacking channels, our approach leads to a class of generalized correlation detectors, which are simple in form, intuitively appealing, and optimal at the same time.

The alternative H_1 is a composite hypothesis with non-Gaussian noise. LRT leads to the critical region $\overline{\Gamma} = \{\mathbf{z}\text{: } l(\mathbf{z}) > T\}$, which depends on $\vec{\gamma}' = (\gamma'_1, \ldots, \gamma'_N)$, and no UMP test is available for this composite hypotheses testing [32]. Without UMP, now we seek an optimum decision rule in the sense of locally most powerful detection (LMP).

For the transformed coefficients of DCT, DWT, or pyramid transforms, $x_i \sim$ GGD $(c_i; 0, \sigma_i)$, whose parameters can be estimated using the moments method, the goodness-of-fit method, or the minimum relative entropy method. Parallel channels of generalized Gaussian distributions are used, which include the global stationary GGD, with c_i and σ_i fixed for all i, as a special case. Under H_0, $y_i \sim$ GGD$(c_i; 0, \sigma_i)$, $z_i \sim$ GGD$(c_i; 0, \sigma_i)$; and under H_1, $y_i \sim$ GGD$(c_i; 0, (1 + \gamma_i w_i)\sigma_i)$, $z_i \sim$ GGD$(c_i; 0, (1 + \gamma'_i w_i)\sigma_i)$.

To conserve the mean value of the transformed coefficients so that there is no "change of lighting conditions" undergone to the watermarked image, zero-mean watermarks are used, which include the CDMA sequence [14], [19] and the random sequence drawn uniformly from $[-1, 1]$. Since γ_i represent the allowable variations of watermark signals relative to the local neighborhood, from Weber's law, γ_i are small, and modulated "physical" watermarks are typical weak signals. Due to the imperceptibility requirement, $\gamma_i < 1$ in watermarking applications, and actually, it is found out that usually, $\gamma_i < 0.3$ in perceptual evaluations (see Section V).

Now, denote

$$F(\vec{\gamma}') = \prod_{i=1}^{N} \frac{1}{(1 + \gamma'_i w_i)}. \tag{13}$$

Then

$$p_1(\mathbf{z}) = A^N F(\vec{\gamma}') \exp\left(- \sum_{i=1}^{N} \frac{|\beta_i z_i|^{c_i}}{(1 + \gamma'_i w_i)^{c_i}} \right) \tag{14}$$

and

$$\frac{d}{d\vec{\gamma}'} F(\vec{\gamma}') = \left(\sum_{i=1}^{N} \frac{-w_i}{1 + \gamma'_i w_i} \right) F(\vec{\gamma}'). \tag{15}$$

Since $\mathbf{w}$ is a zero-mean sequence

$$\frac{1}{N}\ d/d\vec{\gamma}' F(\vec{\gamma}')|_{\vec{\gamma}'=0} = -\frac{1}{N} \sum_{i=1}^{N} w_i \tag{16}$$

which does not depend on $\mathbf{z}$ and tends to 0 as $N \to \infty$ by the weak law of large numbers (WLLN). Thus, $(1/N) p_0^{-1}(\mathbf{z})(d/d\vec{\gamma}') p_1(\mathbf{z})|_{\vec{\gamma}'=0}$ yields the locally optimum decision statistic $((1/N)\sum_{i=1}^{N} w_i c_i \beta_i{}^{c_i} |z_i|^{c_i})$, which is summarized in the following theorem.

Theorem 1: Assume $\mathbf{x} = \{x_1, \ldots, x_N\}$ is a sequence of independent random variables with $x_i \sim$ GGD$(c_i; 0, \sigma_i)$, and $\mathbf{w} = \{w_1, \ldots, w_N\}$ is a known, zero-mean sequence valued in $[-1, 1]$, which is statistically independent of $\mathbf{x}$. For the multiplicative watermarks in (2) with $0 < \gamma_i < 1$, the robust optimum decision statistic is given by

$$\delta(\mathbf{z}) = \frac{1}{N} \sum_{i=1}^{N} c_i |\beta_i z_i|^{c_i} w_i \gtrless \eta \quad \begin{matrix} \Rightarrow \mathbf{w} \text{ exists} \\ \Rightarrow \text{No } \mathbf{w} \end{matrix} \tag{17}$$

where η is a proper threshold, and $\mathbf{z} = \{z_i\}_{i=1}^{N}$ is a sequence of output of the watermarking channel.

The locally most powerful detection is also known as locally optimum detection (LOD) [32], [33]. It is optimum for weak signals [34] and can still perform very well even when the signal strengths become large [35]. From the Pitman–Noether theorem, it can be shown that LOD is the most efficient asymptotically [32], [33]. Since the optimum detectors in the above theorem correlate watermark signals with the observation magnitudes raised to the power of c_i, we call them generalized correlation detectors or generalized correlators. If the sub-band coefficients are modeled using a global stationary GGD$(c; 0, \sigma)$, the decision statistic can be made even simpler with $c\beta^c$ absorbed into the threshold. The value of γ_i can be determined by perceptual evaluations for watermarked images. The larger γ_i under perceptual constraints, the more robust $\delta(\mathbf{z})$ is.

The computational complexities of the LOD and UMP detectors are compared. For each term in the summands in (8) and (17), the computations needed are tabulated in Table I. To further reduce the computational complexity, we demonstrate in Section V that the following detector can be employed as a "*universally optimal*" detector, in the sense that parameters of GGD need not be estimated for each image or video frame any longer (the estimation may pose computational burdens for watermarking applications in DVD, video, or real-time systems),

TABLE I
COMPUTATIONAL COMPLEXITY COMPARISONS OF EACH TERM IN SUMMANDS OF LOD AND UMP DETECTORS, CORRELATOR, AND SRD. COMPUTATIONS ARE ADDITION (ADD.), MULTIPLICATION (MULT.), TAKING POWER (POWER), AND ESTIMATION (EST.) OF GGD PARAMETERS β AND c

	Add.	Mult.	Power	Est. of β and c
SRD	0	1	1	No
LOD	0	2	1	Yes
UMP	2	3	2	Yes
Correlator	0	1	0	No

and near-optimal performance can be obtained for a large set of natural images and video frames:

$$\delta(\mathbf{z}) = \frac{1}{N}\sum_{i=1}^{N} |z_i|^{1/2} w_i \geq \tilde{\eta} \quad \Rightarrow \mathbf{w} \quad \text{exists}$$
$$< \tilde{\eta} \quad \Rightarrow \text{No} \quad \mathbf{w} \tag{18}$$

where $\tilde{\eta}$ is a proper threshold, $\mathbf{z} = \{z_i\}_{i=1}^{N}$ is a sequence of output of the watermarking channel, and $\mathbf{w}$ is a watermark to be verified. Since the square roots of the observation magnitudes are correlated with the watermark, we call it the square-root detector (SRD).

The mean and variance of the decision statistic in (17) can help us to determine a proper threshold η. Under H_0, the mean and the variance of $\delta(\mathbf{z})$ are

$$E(\delta(\mathbf{z}|H_0) = \frac{1}{N}\sum_{i=1}^{N} E(w_i)E(|z_i|^c) = 0 \tag{19}$$

and

$$Var(\delta(\mathbf{z}|H_0) = \frac{1}{N^2}E(|x|^{2c})\sum_{i=1}^{N} E((w_i)^2) \tag{20}$$

where $x \sim \text{GGD}(c; 0, \sigma)$. Since w_i is valued in $[-1, 1]$, $E((w_i)^2)$ is bounded above by 1, and the upper bound is achieved by a CDMA pseudo-noise sequence with $w_i = 1$ or -1 equiprobably. Thus, $(1/N^2)\sum_{i=1}^{N} E((w_i)^2) \leq 1/N$, and $Var(\delta(\mathbf{z})|H_0) \to 0$ as $N \to \infty$, namely, the decision statistic under H_0 asymptotically approaches 0 as N increases.

Let $\overline{\sigma}_w^2 = (1/N)\sum_{i=1}^{N} E(w_i)^2$. It can be shown that

$$\sigma_L^2 := Var(\delta(\mathbf{z}|H_0) = \frac{1}{N}\frac{2A\overline{\sigma}_w^2\Gamma\left(2+\frac{1}{c}\right)}{c\beta^{1+2c}} \tag{21}$$

where $\Gamma(\cdot)$ is the Gamma function. Using Gaussian approximation, under H_0, $\delta(\mathbf{z}) \sim \mathcal{N}(0, \sigma_L^2)$, for any given false alarm rate $p_f^\star$, also known as the significance level, we obtain a threshold

$$T = \sigma_L Q^{-1}(p_f^\star) \tag{22}$$

where $Q^{-1}(\cdot)$ is the inverse of $Q(\cdot)$, which is the tail probability of a unit Gaussian distribution, namely, $Q(t) = (1/\sqrt{2\pi})\int_t^\infty e^{-(u^2/2)}\,du$.

The mean and variance of $\delta(\mathbf{z})$ under H_1 generally do not have explicit forms; however, for special cases with $c = 1$ and $c = 2$, simple expressions for them are shown in following sections.

D. LODs for Laplacian and Gaussian Models

DCT coefficients in AC sub-bands can be reasonably modeled using the Laplacian law [28], [36], which is a GGD law with $c = 1$. The coefficients of sub-band transforms as well as those of the pyramid transform may also be reasonably modeled using the Laplacian law [36]. The pdf of the Laplacian law is

$$p(x) = \frac{1}{\sqrt{2}\sigma}e^{-(\sqrt{2}|x|/\sigma)}. \tag{23}$$

For a global stationary Laplacian model, under the same condition as in Theorem 1, a generalized linear correlator is obtained:

$$\delta(\mathbf{z}) = \frac{1}{N}\sum_{i=1}^{N} w_i|z_i|. \tag{24}$$

To derive the mean and the variance for the Laplacian model, in addition to the conditions in Theorem 1, we assume that the watermark signals are i.i.d. random variables, with $Ew_i = Ew_i^3 = 0$, $Ew_i^2 = \sigma_w^2$, and $Ew_i^4 = \kappa_4$, $i = 1, \ldots, N$. This is true if the pdf of w_i is symmetric about the origin; for example, the CDMA pseudo-noise sequence or the watermark signal drawn from the uniform distribution in $[-1, 1]$ has this property. Then, under H_0

$$E(\delta(\mathbf{z})|H_0) = 0 \tag{25}$$

$$Var(\delta(\mathbf{z})|H_0) = \frac{\sigma^2\sigma_w^2}{N}. \tag{26}$$

Under H_1, using straightforward computations and $E|x| = \sigma/\sqrt{2}$

$$E(\delta(\mathbf{z})|H_1) = \frac{1}{\sqrt{2}}\overline{\gamma}\sigma\sigma_w^2 \tag{27}$$

$$Var(\delta(\mathbf{z})|H_1) = \frac{1}{2N}(2\sigma^2\sigma_w^2 + \overline{\gamma}^2\sigma^2(2\kappa_4 - \sigma_w^4)) \tag{28}$$

where $\overline{\gamma} = (1/N)\sum_{k=1}^{N}\gamma_k$, and $\overline{\gamma}^2 = (1/N)\sum_{k=1}^{N}\gamma_k^2$.

The Gaussian noise model is commonly used in the literature, which is a special GGD law with $c = 2$. For a global stationary Gaussian model, under the same condition as in Theorem 1, a generalized quadratic correlator is obtained:

$$\delta(\mathbf{z}) = \frac{1}{N}\sum_{i=1}^{N} w_i z_i^2. \tag{29}$$

To determine the mean and variance of $\delta(\mathbf{z})$, similar to the approach in the Laplacian case, we assume that the watermark sig-

TABLE II
DESCRIPTION OF THE VISUAL QUALITY OF MULTIPLICATIVELY WATERMARKED IMAGES IN THE DCT DOMAIN WITH DIFFERENT VALUES OF α

Embedding Strength α	Lena	Bridge	Couple
0.08	Identical	Identical	Identical
0.15	Identical	Identical	Identical
0.19	Identical	Identical	Identical
0.20	Identical	Identical	Little distortions at some edges
0.25	Less sharp at some edges	Tiny stains on bridge body	A little distortions at some edges
0.30	Scar around an edge	Some stains on bridge body	More distortions at some edges
0.35	Scar around an edge	More stains on bridge body	More distortions at some edges
0.40	Distortions at plain area obvious near edges	Many stains on bridge body	Distortions spread image

nals are i.i.d. random variables, with $Ew_i = Ew_i^3 = Ew_i^5 = 0$, $Ew_i^2 = \sigma_w^2$, $Ew_i^4 = \kappa_4$, and $Ew_i^6 = \kappa_6$, $i = 1, \ldots, N$ in addition to the conditions in Theorem 1. Then, by some straightforward computation and using $Ex^4 = 3\sigma^4$, we get

$$E(\delta(\mathbf{z})|H_0) = 0 \tag{30}$$

$$Var(\delta(\mathbf{z})|H_0) = \frac{3\sigma_w^2\sigma^4}{N}. \tag{31}$$

Under H_1

$$E(\delta(\mathbf{z})|H_1) = 2\overline{\gamma}\sigma^2\sigma_w^2, \tag{32}$$

$$Var(\delta(\mathbf{z})|H_1) = \frac{1}{N}\left(3\sigma^4(\sigma_w^2 + 6\overline{\gamma}^2\kappa_4 + \overline{\gamma}^4\kappa_6) - 4\overline{\gamma}^2\sigma^4\sigma_w^4\right) \tag{33}$$

where

$$\overline{\gamma} = \frac{1}{N}\sum_{k=1}^{N}\gamma_k, \qquad \overline{\gamma}^2 = \frac{1}{N}\sum_{k=1}^{N}\gamma_k^2$$

and

$$\overline{\gamma}^4 = \frac{1}{N^2}\sum_{k=1}^{N}\sum_{l=1}^{N}\gamma_k^2\gamma_l^2.$$

To maximize the detection probability under a given significance level, namely, to achieve the optimality using the Neyman–Pearson criterion, the threshold given in (22) can be utilized with $c = 2$. Alternatively, for equal priors and uniform cost [37], a threshold may be chosen as $\eta = \overline{\gamma}\sigma^2\sigma_w^2$ to minimize the probability of detection errors.

E. Performance Analysis

In the detection of a specific watermark, there are two kinds of errors: A false alarm occurs when there is no such a watermark but the detector reports its existence, and a miss occurs when there is such a watermark but the detector reports its nonexistence. The detection errors undermine the credibility of the watermarking system and thus need to be controlled strictly. Especially, the false alarms are against the interests of consumers when the detectors are located inside consumer devices for controlling record and playback. Therefore, consumer electronic manufacturers require that the false alarm error ratio be extremely low. Accurate performance analysis of the corresponding watermarking methods is also completely necessary before they can actually be applied in the practice of law enforcement.

Let $u_i = w_i|\beta_i z_i|^{c_i}$, $i = 1, \ldots, N$. Assume $\{x_i\}_{i=1}^N$ are independent and follow a stationary GGD and $\{w_i\}_{i=1}^N$ are i.i.d.; then, $\mathbf{u} = \{u_1, \ldots, u_N\}$ is a sequence of i.i.d. random variables. From the original and the watermarked images, we can estimate the mean under H_θ, $m_\theta = E(u_i|H_\theta)$, and the variance under H_θ, $V_\theta = Var(u_i|H_\theta)$, where $\theta \in \{0, 1\}$. Using the central limit theorem (CLT) [37], $\delta(\mathbf{z}) \overset{H_\theta}{\sim} \mathcal{N}(m_\theta, (V_\theta/N))$. Therefore, the probability of a false alarm is

$$P_f = Q\left(\frac{(\eta - m_0)\sqrt{N}}{\sqrt{V_0}}\right) \tag{34}$$

where $Q(x) = (1/\sqrt{2\pi})\int_x^{+\infty} e^{-(t^2/2)}\,dt$. The probability of a miss is

$$P_m = Q\left(\frac{\sqrt{N}\,(m_1 - \eta)}{\sqrt{V_1}}\right) \tag{35}$$

and the ROC is

$$P_d = Q\left(\frac{\sqrt{V_0}\,Q^{-1}(P_f) + \sqrt{N}\,m_0 - \sqrt{N}\,m_1}{\sqrt{V_1}}\right). \tag{36}$$

Since γ_i is often very small, which implies that $\sqrt{V_0} \approx \sqrt{V_1}$, and m_0 is close to zero for orthogonal transforms, the approximate ROC is

$$P_d = Q\left(Q^{-1}(P_f) - \frac{\sqrt{N}\,m_1}{\sqrt{V_1}}\right). \tag{37}$$

The above performance analysis makes use of CLT, hence Gaussian approximations, to approximate the performance. Since sub-band coefficients are typically non-Gaussian, CLT can only provide crude analysis in this scenario. Future research

TABLE III
DESCRIPTION OF THE VISUAL QUALITY OF MULTIPLICATIVELY WATERMARKED IMAGES IN DWT DOMAIN WITH DIFFERENT VALUES OF α

Embedding Strength α	Lena	Bridge	Couple
0.08	Identical	Identical	Identical
0.12	Identical	Identical	Identical
0.21	Identical	Identical	Identical
0.22	Barely see little shadow at certain edge	Identical	Identical
0.30	Small "burn-out" at some edges	Little ripple on bridge body	Little stain on the door
0.40	Small ripples at some edges	Ripples on bridge body	Stains on the door
0.50	Small ripples replace edges	Texture-like artifacts on bridge body	Texture-like artifacts on the door and the wall

can be done to provide accurate performance analysis beyond CLT, as has been done in [23].

IV. OPTIMUM DETECTION OF MULTIPLICATIVE WATERMARKS IN DFT DOMAIN

Optimum detection for multiplicative watermarks in sub-band transformed domains such as DCT, DWT, and pyramid transform has been constructed. The methodology is extended in this section to the optimum detection of multiplicative watermarks in DFT magnitude domain. For multiplicative watermarks in DFT domain, a decoding rule other than the correlator has been proposed [24], [25]. However, this decoding rule does not take into account the attacks as well as visual masking effects [24]. A new detection rule is constructed for multiplicative watermarks cast in the DFT domain, which is different from the one derived in [24] and [25] in terms of both derivation method and result. The Weibull distribution is adopted, similar to [24], to statistically describe the magnitudes of DFT coefficients. The attacks as well as visual masking effects are incorporated into the formulation.

A. Statistical Model for Magnitudes of DFT Coefficients

The magnitudes of the DFT coefficients are modeled using the Weibull distribution [24], [25]:

$$f_{\mathbf{x}}(x) = \frac{\rho}{\alpha}\left(\frac{x}{\alpha}\right)^{\rho-1}\exp\left[-\left(\frac{x}{\alpha}\right)^{\rho}\right] \tag{38}$$

where α and ρ are positive constants controlling the mean, variance, and shape of the distribution. It is denoted as *Weibull* (ρ, α). The exponential and Rayleigh distributions are special cases with $\rho = 1$ and $\rho = 2$, respectively. The mean μ_X and the standard deviation σ_X are

$$\mu_X = \alpha\Gamma\left(1+\frac{1}{\rho}\right) \tag{39}$$

$$\sigma_X^2 = \alpha^2\Gamma\left(1+\frac{2}{\rho}\right) - \mu_X^2. \tag{40}$$

TABLE IV
DECODING ERRORS FOR BOTH UMP DETECTOR AND LOD DETECTOR AFTER JPEG COMPRESSION WITH DIFFERENT QUALITY FACTORS. SHORT CODEWORD IS USED TO DISCRIMINATE DECODING CAPABILITIES

Quality Factor	100	95	90	85	80	75	70	65	60	55	50	45
UMP	21	21	22	23	21	22	21	22	21	21	20	22
LOD	21	21	22	23	21	22	21	22	21	21	20	22

Using this statistical model, a UMP decoding or detection rule for additive watermarks in the domain of DFT magnitudes can be obtained. The additive embedding rule is

$$y_i = x_i + s_i, \qquad i = 1, \ldots, L \tag{41}$$

where x_i are DFT magnitudes, and s_i are modulated watermark signals. Assuming the decoder knows or can learn from the test signal of the modulated watermark, a simple hypothesis testing can be formed. Assuming $x_i \sim$ Weibull (ρ_i, α_i) and using LRT, an optimum decoder is

$$l_{ad}(\mathbf{y}) = \sum_{i=1}^{L}\frac{y_i^{\rho_i} - (y_i - s_i)^{\rho_i}}{\alpha_i} + (\rho_i - 1)\log\left(1 - \frac{s_i}{y_i}\right). \tag{42}$$

The optimality is in the sense of the Neyman–Pearson lemma. Its derivation and theoretical performance analysis can be done in an analogous way to [23]. Using the additive rule, one constraint is $|s_i| < x_i$ due to the positivity of y_i. Since there is no explicit perceptual model for the DFT magnitudes, while a multiplicative watermark achieves high perceptual quality, we focus on multiplicative watermarking in the DFT magnitude domain.

B. Previous Work

The detection of multiplicative watermarks in the DFT domain is often based on a correlation detector. Recently, the detection of multiplicative watermarks in the DFT domain has been considered with the embedding equation (2), where x_i are the magnitudes of DFT coefficients [24], [25]. Two hypotheses are defined [24]: The system contains a certain watermark $\mathbf{w}^\star = \{w_1^\star, \ldots, w_N^\star\}$ (hypothesis K_1), or the system does not contain

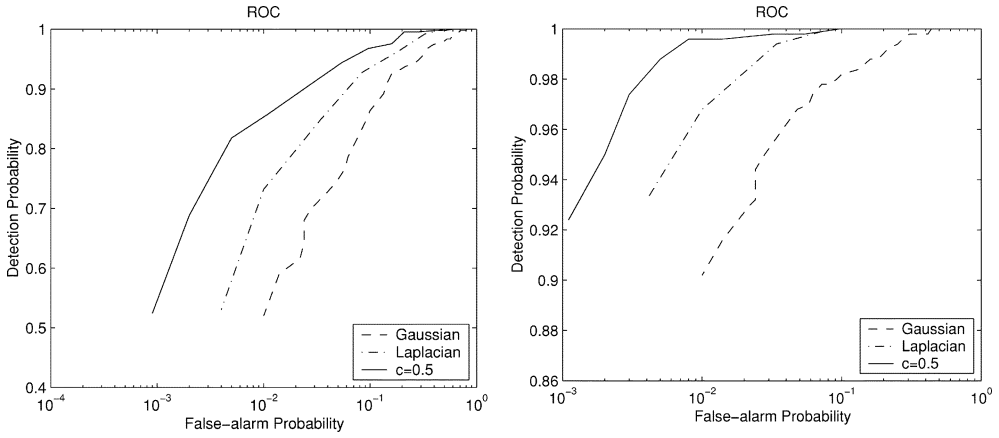


Fig. 1. ROCs for Lena with (a) $\gamma = 0.08$ and (b) $\gamma = 0.12$.

this parameter (hypothesis K_0). The parameter space can be defined as

$$M = M_0 \cup M_1 \tag{43}$$

where

$$M_0 = \{\mathbf{w} \neq \mathbf{w}^\star\}, \qquad M_1 = \{\mathbf{w}^\star\}. \tag{44}$$

The watermark components are assumed to be uniformly distributed in $[-1, 1]$, and M_0 is composed by an infinite number of watermarks. By considering the Weibull distribution of the magnitude of the DFT coefficients, it is demonstrated that $f_{\mathbf{y}}(\mathbf{y}|M_0)$ can be approximated by $f_{\mathbf{y}}(\mathbf{y}|0)$ [24]. Using this simplification, by invoking LRT, a decision statistic is obtained [24]:

$$\sum_{i=1}^{N} \left[\frac{y_i^{\rho_i}[(1+\gamma w_i^\star)^{\rho_i} - 1]}{[\alpha_i(1+\gamma w_i^\star)]^{\rho_i}} \right] > \lambda_1 \tag{45}$$

where ρ_i is the shape parameter for x_i, and λ_1 is a properly chosen threshold. For a given false-alarm rate, a suitable threshold may be determined utilizing CLT.

As noted by the authors in [24], the derivation does not consider attacks as well as visual masking effects. In the embedding stage, a visual masking method is exploited, which modifies the embedding strengths γ pixel-wise [24]. That is, γ_i is different for different x_i. In the derivation of the detector, a single embedding strength $\gamma^\star$ is assumed for employing the Neyman–Pearson lemma [24].

C. Barni's Detector From a Simple Hypothesis Testing

The decoder in [24] can be obtained, actually, through the following simple hypothesis testing: K_1': $\nu = \gamma^\star$ versus K_0': $\nu = 0$ for the same embedding rule: $y_i = x_i + \nu w_i^\star x_i$, $i = 1, \ldots, N$. Indeed, under K_0', y_i has the same distribution as x_i, namely, $f(y_i|K_1') = f(x_i|K_1')$; under K_1', by changing variables, y_i has the following distribution:

$$f(y_i|K_1') = \frac{\rho_i}{\alpha_i(1+\gamma w_i^\star)} \left(\frac{y_i}{\alpha_i(1+\gamma w_i^\star)} \right)^{(\rho_i - 1)} \cdot e^{-(y_i/(\alpha_i(1+\gamma w_i^\star)))^{\rho_i}}. \tag{46}$$

Assuming that the observations $\{y_i\}_{i=1}^N$ are independent under both hypotheses, we have $f(\mathbf{y}|K_\theta') = \prod_{i=1}^N f(y_i|K_\theta')$, where $\theta \in \{0, 1\}$. Invoking LRT [32] and taking the logarithm, Barni's detector in (45) can be obtained.

As shown above, Barni's detector is literally obtained from testing K_1' versus K_0', both of which are simple hypotheses. Since visual masking changes the value of $\gamma^\star$ in casting watermark $\{w_i^\star\}_{i=1}^N$, different γ_is (derived from JND) may be used for different x_is. The attacks also change the values of γ_i. Testing a single value of $\gamma^\star$ neglects the modeling of these effects.

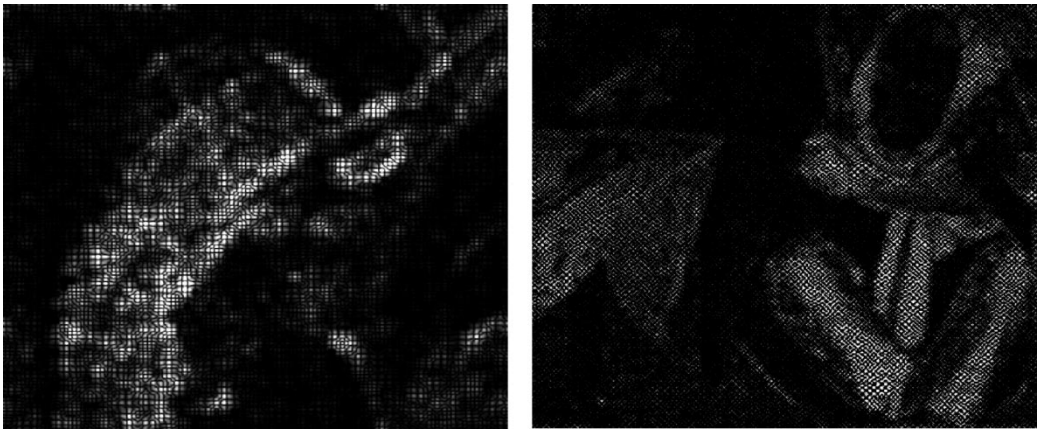
D. Robust Optimum Detection of Multiplicative DFT Domain Watermarks

In this section, robust optimum detectors for multiplicative watermarks in DFT domain are derived. Considering the visual masking and attacking effects, as discussed in Section III, the following hypothesis testing is conducted: $H_1^\star$: $\nu_i' > 0$ versus $H_0^\star$: $\nu_i' = 0$ for $z_i = x_i(1 + \nu_i' w_i^\star)$, which is a composite hypothesis versus a simple one.

Due to the invisibility constraint, ν_i is small. Watermark $\mathbf{w}^\star$ is required to be zero mean to guarantee that there is no "change of lighting conditions" to the watermarked image. CDMA pseudo random sequence [14], [19] is a special case of these watermarks. By adopting LOD, the following result for the detection of the multiplicative watermark in DFT domain is obtained.

Theorem 2: Assume $\mathbf{x} = \{x_1, \ldots, x_N\}$ is a sequence of independent random variables with x_i obeying $Weibull(\rho_i, \alpha_i)$, and $\mathbf{w}^\star = \{w_1^\star, \ldots, w_N^\star\}$ is a known, zero-mean sequence valued in $[-1, 1]$, which is statistically independent of $\mathbf{x}$. At the transmitter end, the multiplicative watermarks are embedded using $y_i = x_i(1 + \nu_i w_i^\star)$ with $0 < \nu_i < 1$. The watermarked image is then sent through the watermarking channel, which introduces distortions to the composite image. At the receiver end, for the output sequence $\mathbf{z} = \{z_i\}_{i=1}^N$, the optimum decision rule is given by

$$\delta(\mathbf{z}) = \sum_{i=1}^{N} \rho_i \left(\frac{z_i}{\alpha_i} \right)^{\rho_i} w_i^\star \gtrless \eta \quad \Rightarrow \mathbf{w}^\star \quad < \eta \quad \Rightarrow \text{no} \quad \mathbf{w}^\star \tag{47}$$



(a) (b)

Fig. 2. Difference images between (a) watermarked Lena and the original and (b) watermarked Barbara and the original (magnitudes amplified by a factor of 100 for viewing purpose).

where η is a properly selected threshold.

Let $v_i := \rho_i(z_i/\alpha_i)^{\rho_i} w_i^\star$, and the sum of v_i is defined as h for convenience, namely, $h = \delta(\mathbf{z}) = \sum_{i=1}^N v_i$. Denote the pdf of h under $H_0^\star$ as $f_h(h|H_0^\star)$. For a given false alarm rate $p_f^\star$, the value of η can be determined from the decision rule (47) by solving

$$p_f^\star = \int_\eta^{+\infty} f_h(h|H_0^\star)\, dh. \tag{48}$$

To figure out pdf $f_h(h|H_0^\star)$, we will first find out the pdf of v_i under $H_0^\star$. Under such a circumstance, $y_i = x_i$, and z_i will be a noisy version of x_i. To proceed without assuming any fixed distribution of the attacks, we consider $v_i = g(x_i)$, and under $H_0^\star$, $g(x_i) = \rho_i(x_i/\alpha_i)^{\rho_i} w_i^\star$. Since $g(x_i)$ is a monotonically increasing or decreasing function of x_i according to the sign of $w_i^\star$, the pdf of v_i under $H_0^\star$ can be evaluated by the following expression [24], [38]:

$$f_{v_i}(v_i|H_0^\star) = \frac{f_{x_i}(x_i)}{|g'(x_i)|}\, u(v_i) \tag{49}$$

where $v_i = g(x_i)$, $g'(\cdot)$ is the first derivative of $g(\cdot)$, and $u(\cdot)$ is the Heaviside function. Therefore, we have

$$f_{v_i}(v_i|H_0^\star) = \frac{1}{\rho_i w_i^\star} \exp\left(-\frac{v_i}{\rho_i w_i^\star}\right) u(v_i). \tag{50}$$

It can be seen that v_i is distributed as an exponential pdf with mean $\rho_i w_i^\star$ and variance $(\rho_i w_i^\star)^2$. It follows from CLT that $h = \sum_{i=1}^N v_i$ under $H_0^\star$ can be assumed to have a normal distribution, with mean and variance given by

$$\mu_h = \sum_{i=1}^N \rho_i w_i^\star \tag{51}$$

$$\sigma_h^2 = \sum_{i=1}^N (\rho_i w_i^\star)^2. \tag{52}$$

Therefore, the false-alarm probability is

$$p_f^\star = Q\left(\frac{\eta - \mu_h}{\sigma_h}\right) \tag{53}$$

where $Q(x)$ is defined as $Q(x) = (1/\sqrt{2\pi}) \int_x^{+\infty} \exp(-(t^2/2))\, dt$. Therefore, for a given false-alarm rate, we have

$$\eta = \mu_h + \sigma_h Q^{-1}(p_f^\star). \tag{54}$$

In particular, given $p_f^\star = 10^{-6}$, we have

$$\eta = \mu_h + 4.75\sigma_h. \tag{55}$$

V. Experiments

Experiments are conducted to demonstrate the effectiveness of the new detection structures constructed in the paper. Multiplicative watermarks are embedded in both DCT and DWT domains. For DWT domain embedding, a three-level DWT using Daubechies' linear-phase 9/7 biorthogonal filterbanks is used. The watermarks are cast in HH sub-bands at Level 3, which are CDMA pseudo-noise sequences with $w_i = 1$ or -1. First, the perceptual quality in multiplicative watermarking is evaluated. For embedding depth α below 0.3, a small stepsize equal to 0.01 is taken to evaluate the perceptual quality; for α above 0.3, a stepsize 0.05 is taken to judge the severity of perceptual degradations. The subjective evaluations have been conducted by two young viewers with normal HVS under normal viewing conditions for images displayed on computer screens. The results are tabulated in Table II for DCT domain embedding and in Table III for DWT domain embedding. It can be seen that DWT domain watermarking has better perceptual quality than in the DCT domain. In the DCT domain, when α is below 0.20, good fidelity can be obtained; in the DWT domain, when α is below about 0.22, excellent imperceptibility is obtained. Even as α becomes as large as above 0.30, perceptual distortions are not so annoying; especially on plain, featureless areas, visual distortions are not as severe as those of additive watermarking.

To evaluate the proposed LOD detector in (8), it is compared to the UMP detector in (17). A fixed embedding strength is used and multiple-symbol messages are encoded into watermarks in a way similar to CDMA embedding [26], [27] by dividing the image into nonoverlapping embedding regions. Each codeword has 32 bits, and one message bit is encoded into one codeword

using binary phase shift keying (BPSK) [26]. The length of the codeword is short and many decoding errors are incurred, which discriminates well the decoding capabilities of both detectors. The decoding errors are tabulated in Table IV. It can be seen that LOD has the same decoding capabilities as UMP.

To compare the performances of LOD detectors, we have fixed some shape parameters for the GGD model in our experiments. The detectors used are generalized quadratic correlator, generalized linear correlator, and SRD.

Lena is watermarked using $\gamma = 0.08$ and $\gamma = 0.12$. Such embedding strengths are used here so that ROC curves can be used to compare the performance. With larger α, ROC curves are all straight lines with detection probabilities all equal to 1. The PSNRs for the watermarked images are 57.99 and 54.47 dB, respectively. The ROC curves are plotted in Fig. 1. It can be seen that SRD outperforms the other two since the shape parameter of HH sub-band at Level 3 of Lena is 0.47, which is close to 0.5.

Then, the robustness of the detectors is tested using extensive experiments. The embedding depth is $\alpha = 0.2$. The differences between the watermarked and original images are displayed in Fig. 2, with magnitudes magnified by 100 for viewing purposes. The watermarked images are compressed using JPEG with different quality factors from 100 to 1, and the compressed images are tested using the proposed detectors. The resulting decision statistics are plotted in Fig. 3. Without watermarking, the decision statistics are 49.99, 0.43, and 0.029, respectively, for $c = 2$, 1, and $1/2$. JPEG compressed watermarked images can be detected.

The ROC curves are plotted for $c = 2$, $c = 1$, and $c = 1/2$ after severe attacks for a large set of images (over 3000 natural images in our database with size 128×128). For example, Fig. 4 plots the ROCs for JPEG compression with quality factor 10 and 5. Figs. 5–7 plot the ROCs for median, average, and Wiener filtering, respectively, with window sizes 3×3 and 7×7. Fig. 8 plots the sample images of Lena after intensity scaling, Gaussian noise adding, and salt-and-pepper noise adding, respectively. Fig. 9 plots the ROCs after these attacks. It can be seen that after severe attacks the SRD still has the best performance since the shape parameters for natural images are usually in the range of $[0.4, 1]$ [22], [39], and many around 0.5. The generalized correlator corresponding to $c = 1/2$ (SRD) can be employed as a fixed robust detector for good overall performance for a large set of images.

The performance analysis derived in Section III-E is also investigated. Experiments are conducted using $\gamma = 0.12$. The empirical ROCs for $c = 2$, $c = 1$, and $c = 1/2$ are obtained by varying the threshold η and the key to the pseudo-random sequence. They are plotted in Fig. 10. The theoretical ROCs are also plotted. It can be seen that the empirical and theoretical performances are in good agreement if the sub-band coefficients are modeled using the Gaussian model. If the Laplacian model or GGD model with $c = 1/2$ is used, then the theoretical performance analysis using (36) is more conservative than the empirical performance. This is because the actual shape parameter for middle- or low-frequency sub-bands is usually in the range of $[0.4, 1]$ [22], [23], whereas CLT is employed for the theoretical analysis, which is crude for non-Gaussian random variables. More accurate performance analysis is a line of future research.

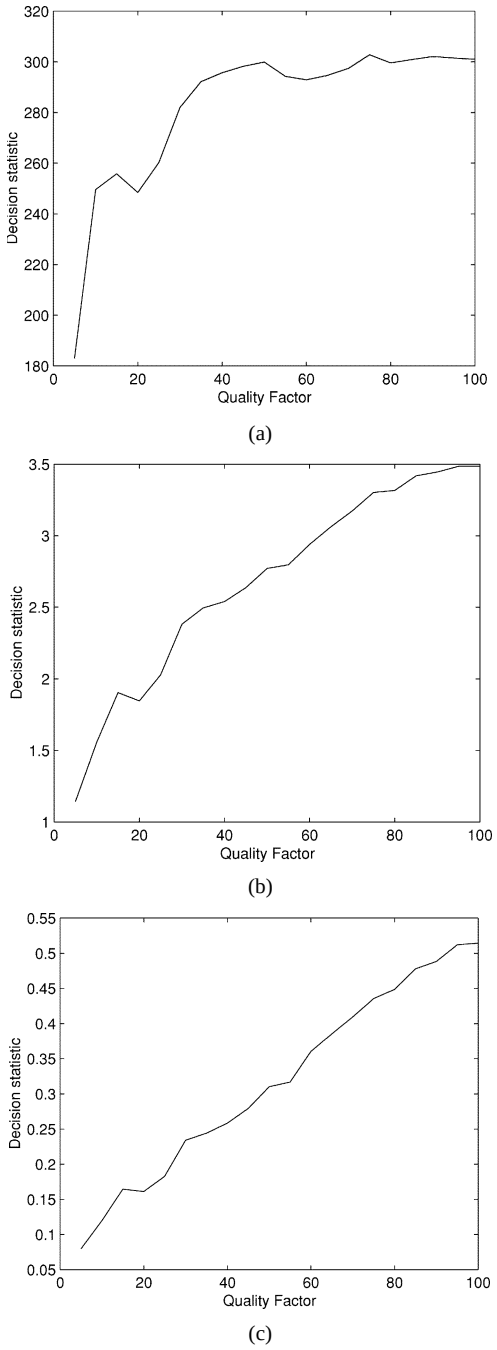


Fig. 3. (a) Decision statistics after JPEG compression with $c = 2$, (b) with $c = 1$, and (c) with $c = 1/2$.

Watermarks are embedded into the DFT domain by modifying a set of full-frame DFT coefficients of the image. The watermarks consist of 1000 length-N bipolar zero-mean pseudo random sequences. One of these sequences is chosen and embedded adopting the multiplicative embedding rule. A visual mask similar to that in [24] is utilized. The watermark is cast into the middle-frequency region of the image. Since the magnitudes of the DFT coefficients are symmetric about the center, the watermark is cast to half the DFT image and mirrors the watermark to the other half. Without loss of generality, it is assumed that $\mathbf{x} = \{x_i\}_{i=1}^{N}$ is drawn from a single Weibull distribution. By using maximum likelihood estimation, the original data are fit to Weibull distribution. For Lena, the estimated exponent is

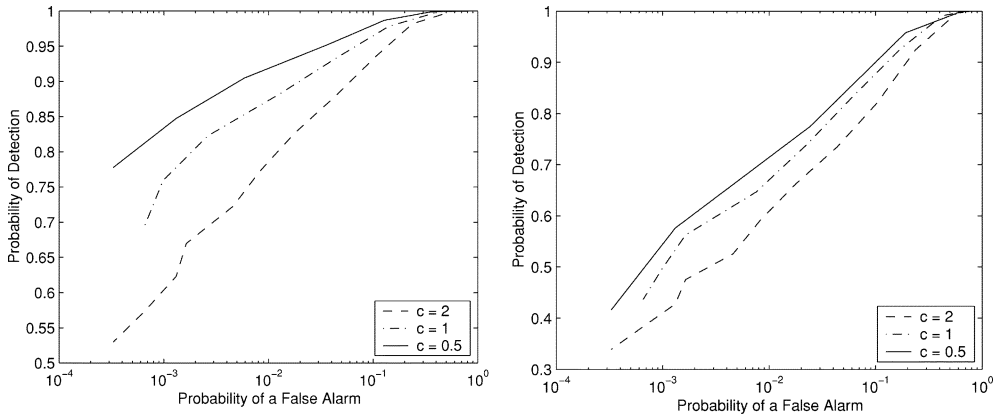


Fig. 4. ROCs for a large set of images after JPEG compression with (a) quality factor 10 and (b) quality factor 5.

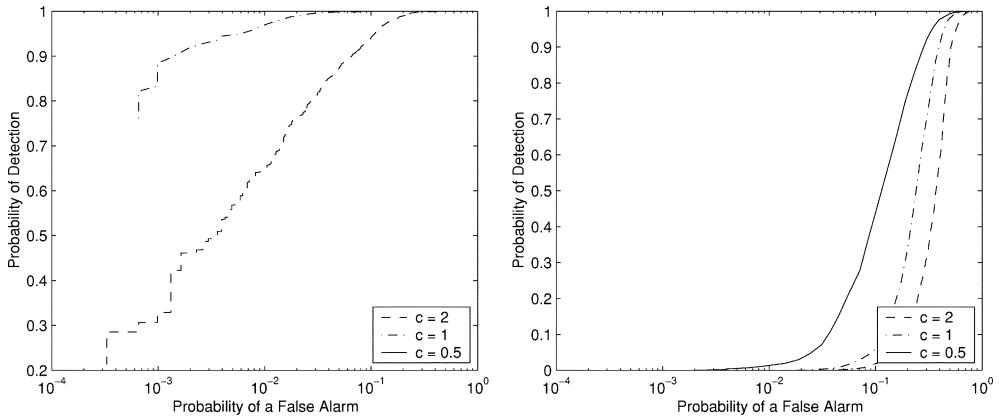


Fig. 5. ROCs for a large set of images after median filtering with size of (a) 3×3 and (b) 7×7.

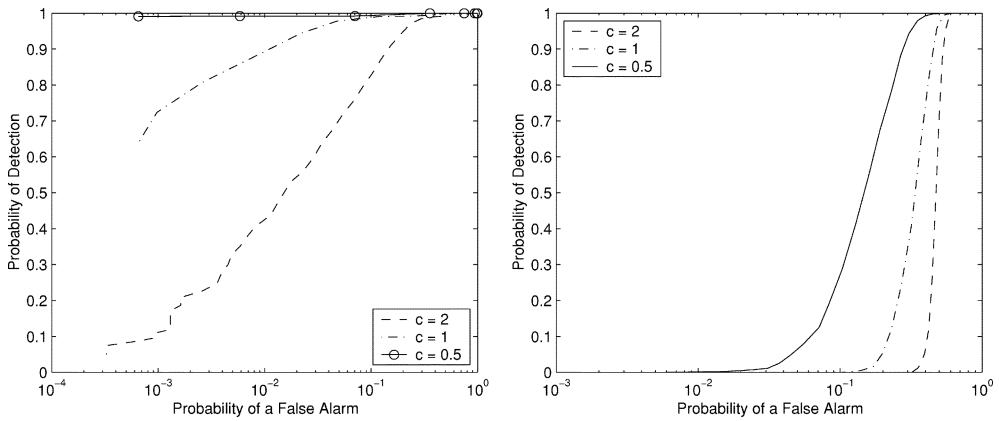
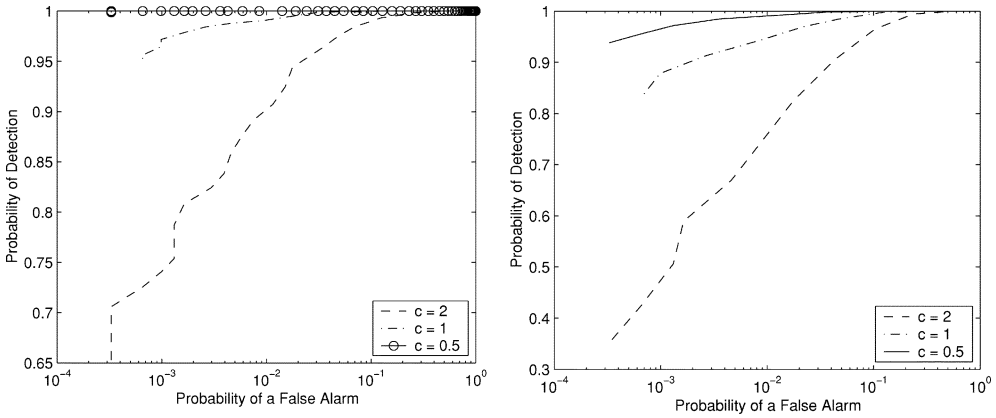


Fig. 6. ROCs for a large set of images after average filtering with size of (a) 3×3 and (b) 7×7.

$\rho = 1.45$. The 500th watermark is embedded as $\mathbf{w}^{\star}$, and $\delta(\mathbf{z})$ of our new detection scheme are plotted in Fig. 11. It shows that the decision statistic differentiates well the embedded one from the rest. After attacking, $\delta(\mathbf{z})$ still performs very well. For example, the decision statistics after the JPEG compression with quality factor 5, and those after median filtering with window size 7×7 are plotted in Fig. 12. The performance of the new detectors is also compared with that of the correlator after attacks. It can be seen from Fig. 13 that the new detectors outperform the correlator after JPEG compression with $QF = 5$.

The experiments have demonstrated the power of the proposed robust optimum detectors. In the following sections, an application of the new detection scheme to combined audio and video watermarking is proposed.



(a) (b)

Fig. 7. ROCs for a large set of images after Wiener filtering with size of (a) 3×3 and (b) 7×7.

VI. Application to Combined Audio and Video Watermarking

The robust optimum detection method can be applied to many watermarking systems. In this paper, we particularly focus on an application to combined audio and video watermarking.

A. Combined Audio and Video Watermarking for Authentication

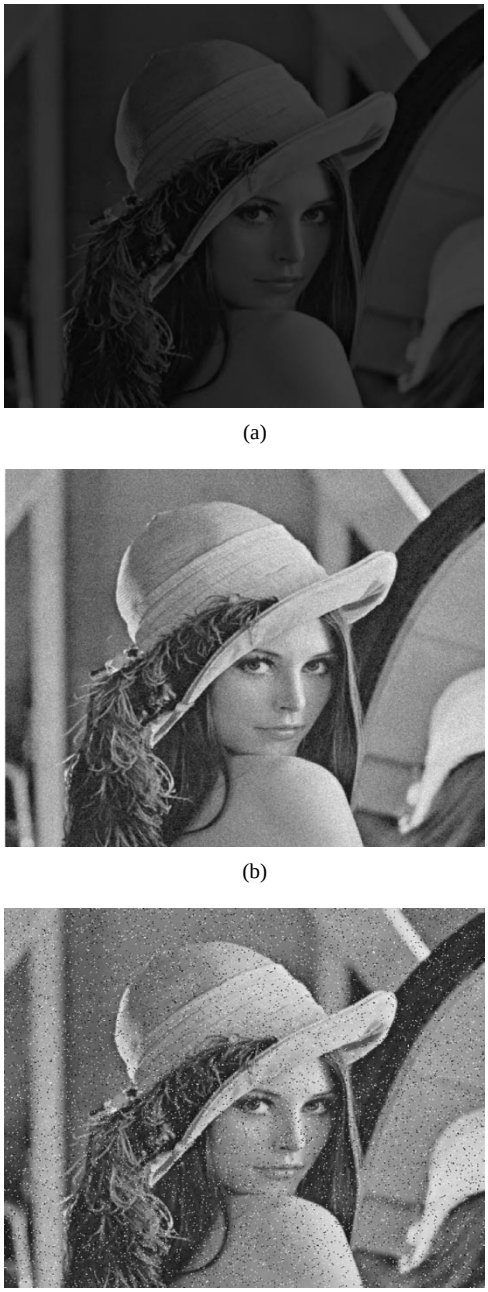
Digital watermarking includes the robust watermarking to notify or protect the copyrights and the fragile watermarking to authenticate the multimedia content. The fragile watermarking modifies the host signal in a way such that the modification can hardly be detected by human eyes or ears, but with a specifically designed detector, it can reliably test the authenticity of the image or even localize the tampered regions.

The fragile watermarking for multimedia authentication is different from the classical cryptographic authentication such as message authentication code (MAC). It does not need a separate file to append to the original one, and it is capable of verifying only the content instead of the data, which allows more tolerance to commonly used data manipulations in transmission and storage. The basic requirements for fragile watermarking are invisibility, fragility, and reliability: The watermarks should be transparent to human eyes or ears; they should be fragile to content editions or alterations, and the verification scheme should make a vanishingly small number of errors since the verification results may be used in the legal process. The multimedia authentication systems may be of interest to any parties who want to tamperproof the content of the multimedia so that they can ensure that the content has not been edited, altered, or damaged. These systems have applications in commerce, defense, journalism, and law.

In the literature, most research in authentication has focused only on single media like video or audio, but the typical multimedia stream consists of both video and audio data, and watermarking only one medium might not guarantee the integrity of the multimedia data. For example, in journalism or teleconferencing, it is quite possible that the watermarked video may not be altered, but its accompanying audio could be edited. In this case, the watermark embedded in the video is not disturbed; however, the content or the semantics of the audiovisual data may be totally changed. Therefore, it is necessary to protect against alteration of one of the components without interfering with the other. Watermarking both components may provide a solution to this problem. However, if the watermarks for the components are independent, the copy attack [40] may be used to copy a watermark from a legitimately watermarked material to another without knowing the embedding system and cryptographic keys. Since audiovisual data may undergo several post production operations such as cutting, editing, and setting in other audiovisual data, the integrity and synchronization of both components in each segment are more important than the integrity of one component alone. An approach to countering these attacks and ensuring integrity and synchronization is to watermark the combined audio and video, where a cryptographic signature is linked to the multimedia content. In the literature, Dittmann *et al.* [41] proposed two solutions for the protection of combined audio and video data. The first one is to insert the same time code to both audio and video data. The audio and video data are not combined, and copy attack can be used to make unauthorized embedding of watermarks. The second one uses the changing of the signs of the audio sequence as an audio feature, which is encoded into the video as the video watermark; it uses the Canny edge detector extracting the edge information of a video frame and embeds the feature code into the audio as the audio watermark. The audio feature is low-level and not very robust to audio compressions. We exploit the characteristics of the audiovisual data and propose a new combined video and audio watermarking method. The mel-frequency cepstra are employed to extract the content information from audio. The feature codes of audio are mapped to watermark patterns cryptographically and imprinted using multiplicative watermarking into video. The optimum detector for multiplicative watermarking is applied. The scheme is tolerable to common video and audio compressions but sensitive to content changes.

B. Information Extraction From Audio

Our goal is to embed the information of one media into the other to verify the integrity of audiovisual data. The human auditory system (HAS) is much more sensitive to noise than the



(a) (b) (c)

Fig. 8. Lena, a sample image from a large set of images. (a) After intensity scaling with scaling factor 0.3. (b) After Gaussian noising with variance 0.002. (c) After adding salt-and-pepper noise with intensity 0.05.

human visual system (HVS). Powerful audio-processing freeware is available on the Internet, which can change the structure of audio signals dramatically while retaining acceptable subjective quality. With these difficulties in mind, we choose to extract audio subjective-content information and embed it into the video.

The well-known technique of mel-frequency cepstra has proven to be highly effective in automatic speech recognition and in modeling the subjective pitch and frequency content of audio signals. Psychophysical studies have found the phenomena of the mel pitch scale and the critical band, and the frequency scale-warping to the mel or Bark scale has led to the cepstrum domain representation. The warped cepstral distance has been defined accordingly [42]:

$$\tilde{d}_2^2(S, S') = \int_{-B}^{B} |\log S(\theta(b)) - \log S'(\theta(b))|^2 \frac{db}{2B} \quad (56)$$

where b is frequency in Barks, $S(\theta(b))$ is the spectrum on a Bark scale, B is the Nyquist frequency in Barks, and θ is a nonlinear function mapping the Bark scale b to the linear scale ω. If two audio signals have similar subjective contents, the warped cepstrum distance should be small.

To incorporate the critical band phenomena into the cepstrum domain representation and to simulate the subjective spectrum efficiently, a filterbank spaced uniformly on the mel scale is used. The modified spectrum of $S(\omega)$ consists of the output power of these filters when $S(\omega)$ is the input. Denoting these power coefficients by $\tilde{S}_k$, $k = 1, 2, \ldots, K$, the mel-frequency cepstrum coefficients (MFCC) $\tilde{c}_n$ are defined as [43]

$$\tilde{c}_n = \sum_{k=1}^{K} (\log \tilde{S}_k) \cos\left[n\left(k - \frac{1}{2}\right)\frac{\pi}{K}\right], \qquad n = 1, 2, \ldots, L \quad (57)$$

where L is the length of the cepstrum, and K is the number of filters in the filterbank.

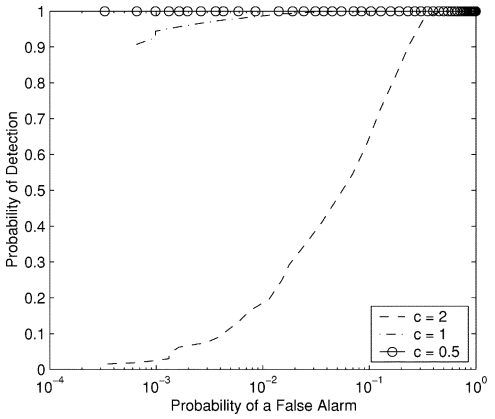
Since the audio signal might undergo many manipulations, the extracted features need to be robust against all allowed attacks, such as filtering, compression, or MP3 encoding. As long as the content has not been modified, the extracted features need to be similar. Toward this end, we choose to make use of the first MFCC of the truncated audio signals with a sliding window, for its well capturing the envelope of the audio signals, and for its robustness. The original audio signals are first downsampled to a low sampling rate; then, the filterbank is applied to compute the MFCC. Fig. 14 shows the first MFCC before and after the IMA ADPCM compression with bandwidth 4 kHz. The original speech has bandwidth 22.05 kHz. The downsampling factor is 4. It can be seen that the shapes are approximately the same.

C. Embedding Feature Code Into Video

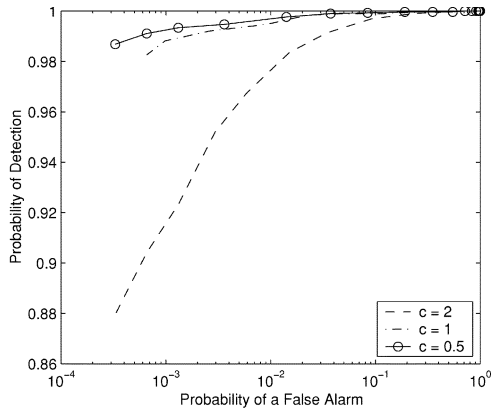
The audio features are extracted by using MFCC so that they are audio content-dependent but at the same time insensitive to manipulations. To further reduce the sensitivity, we make use of the first MFCC coefficient only and quantize the centralized MFCC to several crude levels.

The audio MFCC frame rate is chosen so that each video frame corresponds to several MFCCs. These MFCCs are mapped using a secret function to an integer, which is used as the cryptographic key to generate a bipolar pseudorandom noise pattern $\mathbf{w} = \{w_1, \ldots, w_N\}$. It is the watermark to be embedded into the video. For a simple example, the number of MFCCs could be chosen as 2 for each frame, and the quantization level is chosen as 4; then, there are 16 possible bi-tuple patterns. The encoding function is a one-to-one bijection between 16 bi-tuple vector patterns and 16 pseudorandom noise patterns, which are mutually orthogonal.

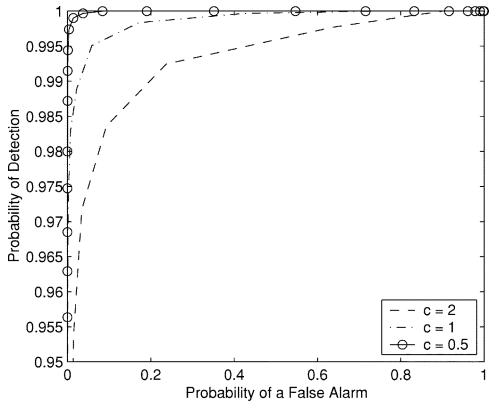
Many schemes have been proposed for video watermarking [16], [44]–[47]. A joint watermarking and compression of the



(a)



(b)



(c)

Fig. 9. ROCs for a large set of images after (a) scaling with scaling factor 0.3. (b) Gaussian noise adding with variance 0.002. (c) Salt-and-pepper noise adding with intensity 0.05.

video scheme has been proposed [46]. The amount of compression of the most recently coded frame of the same type as the current frame was monitored, and the watermark added was adapted according to the degradation due to coding. In this paper, we only adapt the watermark strengths to the frame types. We adopt a multiplicative watermarking rule in the DCT domain:

$$y_k^{(n)} = x_k^{(n)}\left(1 + \alpha_n w_k^{(n)}\right), \qquad k = 1, \ldots, N \tag{58}$$

where $\mathbf{x}^{(n)} = \{x_1^{(n)}, \ldots, x_N^{(n)}\}$ is a series of DCT coefficients of the nth video frame, $\mathbf{w}^{(n)} = \{w_1^{(n)}, \ldots, w_N^{(n)}\}$ is the bipolar

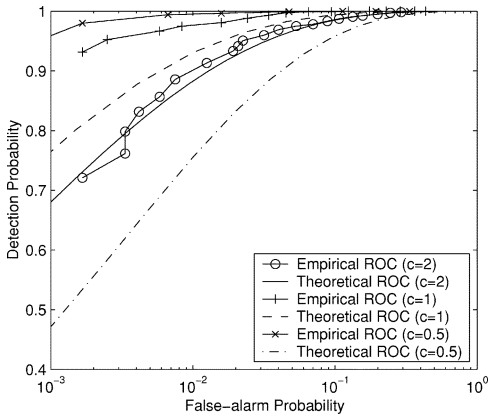


Fig. 10. Empirical ROCs and theoretical ROCs with $\gamma = 0.12$.

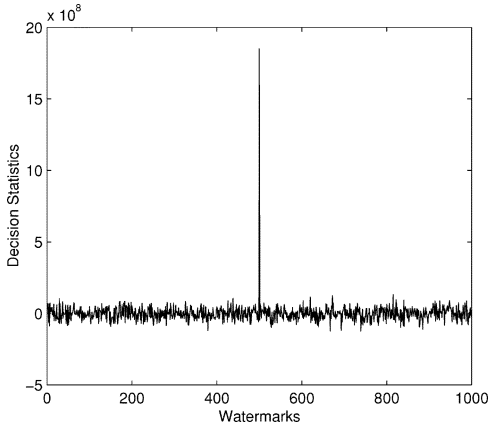
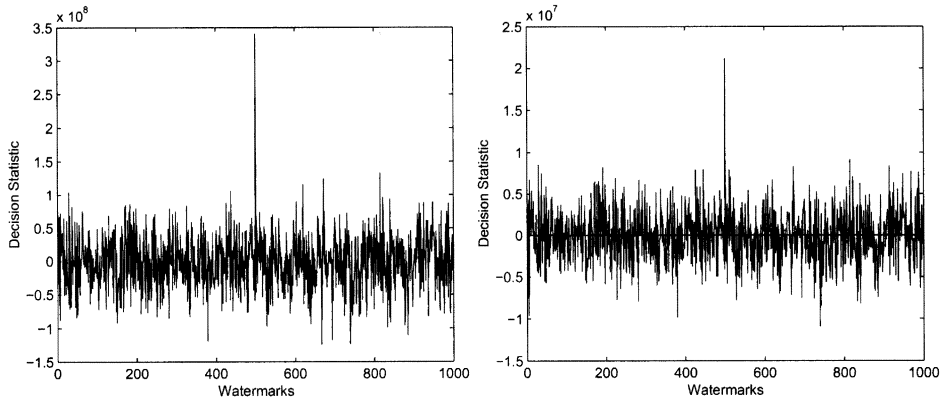


Fig. 11. Decision statistic for the watermarks embedded into the DFT domain. The truly embedded one (the 500th of 1000) produces a very high value compared with the others.

watermark formed from the audio features, and α_n is the watermark strength. Since the I frames are intracoded using DCT only, we set corresponding α_n to a small value, which roughly represents the Weber's fraction [19], [48]. The P frames are forward predicted using the motion-compensated prediction, and the residual errors are compressed using DCT. The degradation due to compression is more than that of I frames. The corresponding α_n is set larger. The B frames are coded most aggressively, and the degradation due to coding is the largest. Moreover, the watermarks embedded in the B frames do not propagate. The corresponding α_n is the largest.

D. Verification of Audiovisual Data

The features of the accompanying audio are first extracted for the test audiovisual data. The extraction is similar to the embedding process, and the parameters such as the audio frame rate, the quantization level, and the encoding function are exactly the same. The extracted features are encoded and mapped to a bipolar pseudorandom sequence $\tilde{\mathbf{w}}^{(n)}$. If there is no modification to the original accompanying audio data at all, $\tilde{\mathbf{w}}^{(n)} = \mathbf{w}^{(n)}$. If there is a certain digital data processing, $\tilde{\mathbf{w}}^{(n)} \approx \mathbf{w}^{(n)}$. However, if the audio data have been edited or falsified, there will be little similarity between $\tilde{\mathbf{w}}^{(n)}$ and $\mathbf{w}^{(n)}$, which can be measured using the Hamming distance. To guarantee the mutual verification, we need to detect the watermark embedded in



(a) (b)

Fig. 12. Decision statistic for watermarked image after attacks. The watermark is embedded into the DFT domain. (a) JPEG compression with quality factor 5. (b) Median filtering with window size 7×7. The truly embedded one (the 500th of 1000) is detected against the others.

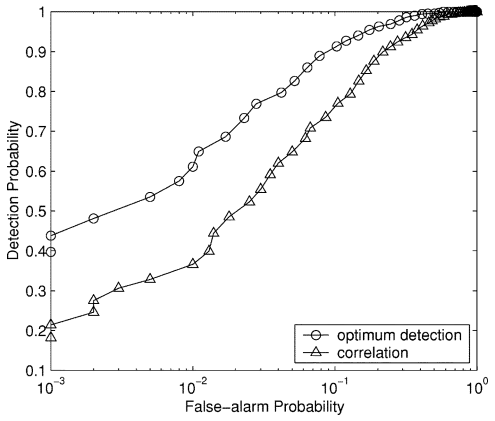


Fig. 13. Comparison of ROC of the new detector after JPEG compression with quality factor 5 to that of the correlator. The watermark is embedded into the DFT domain.

the video data. The decision statistic for the watermarking is defined as the generalized quadratic correlator:

$$\delta\left(\mathbf{y}^{(n)}\right) = \frac{1}{N}\sum_{k=1}^{N}\left|y_k^{(n)}\right|^{1/2}\tilde{w}_k^{(n)}. \tag{59}$$

Fig. 15 shows the decision statistic for a segment of Flower Garden with accompanying audio.

For L GOPs of audiovisual data, we define the authenticity indicator as $AI = u((1/LN_g)\sum_{n=1}^{LN_g}\delta(\mathbf{y}^{(n)}) \geq \eta_0)$ to indicate the authenticity of the content, where η_0 is a properly chosen threshold, N_g is the number of frames in one GOP, and $u(\cdot)$ is the Heaviside function. The binary value of AI indicates the authenticity or nonauthenticity of this GOP.

The video watermarking needs to be tolerant to certain degrees of video compression, such as MPEG-1. The robustness of a video watermarking scheme to MPEG first depends on the chosen bitrate and second on the variation of the embedded signals over consecutive frames. It has been proposed that the watermark be the same for at least the length of a complete GOP [49]. Since the GOP structure might be changed after reencoding, here, we choose to generate different watermarks for different frames.

To validate our proposed combined audio and video watermarking system, we have conducted experiments on the Flower Garden sequence accompanied with a segment of audio signal whose sampling rate is 44.1 Hz. The features extracted from the audio data are the first coefficients of the MFCC with the audio frame rate 60 Hz. For I frames, the watermarks are modulated with $\alpha_n = 0.16$, which is the same as that in [46]. For P frames, α_n is taken as 0.18. For B frames, we take $\alpha_n = 0.22$. Here, we adapt α_n simply to the types of frames. The DCT-domain adaptive scheme proposed in [46] may be incorporated into our framework to further improve the detection results. We use $\eta_0 = 0.15$ and the number of GOPs $L = 4$ in our experiments.

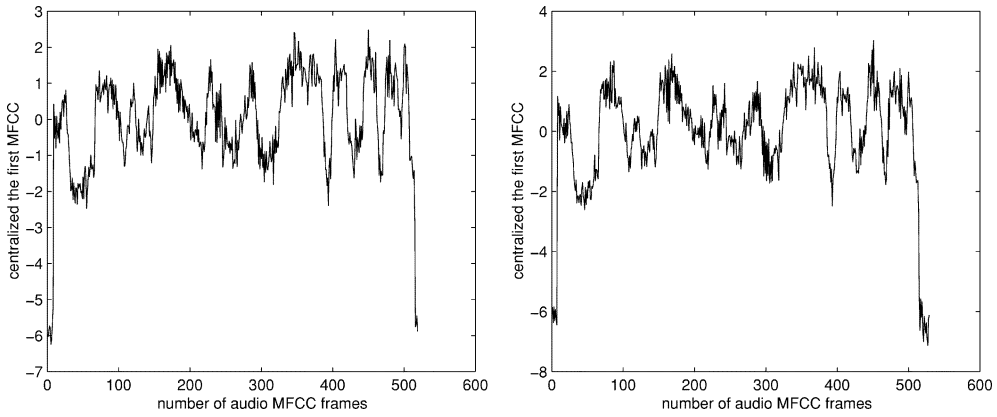
Subjective evaluations of watermarked video have shown no appreciable perceptual quality loss incurred due to watermarking. For 48 frames, the results of $\delta(\mathbf{y}^{(n)})$ for both watermarked and unwatermarked video are shown in Fig. 15. The corresponding authenticity indicator is 1.

To show that watermarks are persistent after video compression, we use MPEG-1 to compress the watermarked video sequence. Fig. 16 shows the resulting $\delta(\mathbf{y}^{(n)})$ for compressed video with data rate 1.152 Mb/s, which corresponds roughly to the VHS quality. The authenticity indicator is 1.

To show the effectiveness of the proposed scheme after audio compression, we use IMA ADPCM to compress the audio signal with bandwidth 4 kHz. This aggressive compression is often used in speech coding with telephone quality. Fig. 17 shows the resulting decision statistics. The results are exactly the same as those without any audio compression.

For a different segment of the same audio data, which is used to falsify the original audio data, the results of the decision statistic are shown in Fig. 18. The authentication indicator is 0 in this case.

The proposed combined audio and video watermarking method is also compared with that in [41], which extracts changing of signs as audio features to embed into video, and extracts edge information as video feature to embed into audio. For comparison, the results of the decision statistic watermarked video are shown in Fig. 19, with audio and video



(a) (b)

Fig. 14. (a) $\bar{c}_1$ for the original speech signal. (b) $\bar{c}_1$ for compressed speech using IMA ADPCM with bandwidth 4 kb/s.

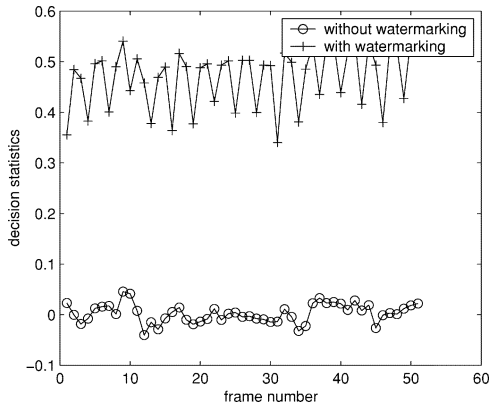


Fig. 15. Decision statistic for audiovisual data without and with watermarking.

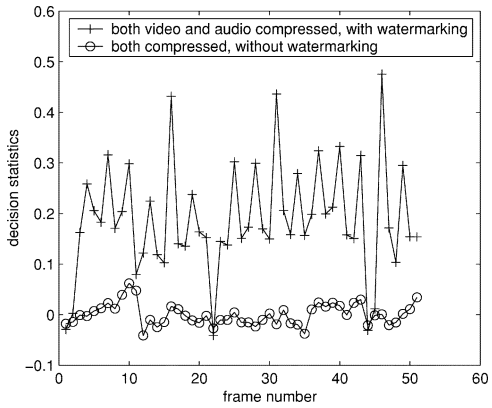


Fig. 17. Decision statistic for compressed audio only, and for compressed audio and compressed video.

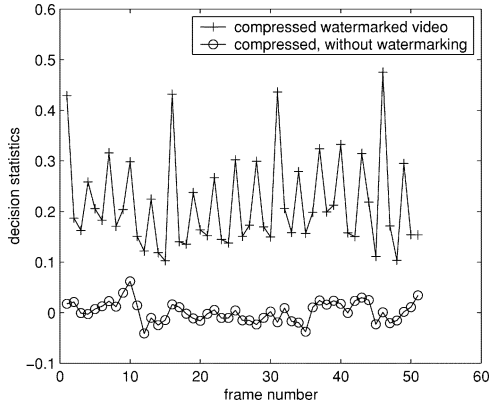


Fig. 16. Decision statistic for compressed watermarked video data.

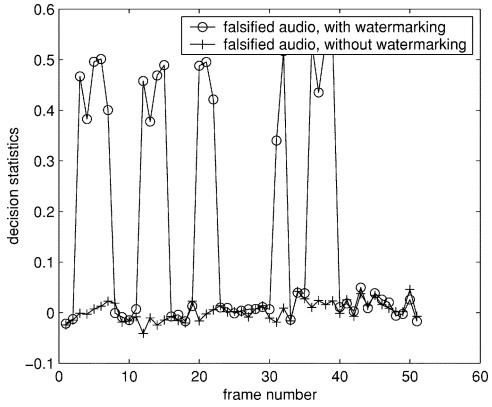


Fig. 18. Decision statistic for falsified audio.

both compressed. Compared with Fig. 17, better results are obtained using the proposed method.

VII. Conclusion and Future Research

In this paper, novel optimum detectors for multiplicative watermarks are derived using locally optimum detection for the generalized Gaussian distributions. Special cases for the Laplacian model and the Gaussian model have been examined. The square-root detector has been shown to be nearly optimum for majority natural images. Novel optimum detectors for the Weibull distribution have also been constructed. The experiments have validated the theoretical analysis. The results can be applied to multiplicative watermarking systems for copyright notification and protection and fingerprinting. It may also find applications in content-based multimedia indexing and retrieval. We have applied the robust watermarking detection to combined audio and video watermarking. It can tolerate commonly used audio and video compressions but is sensitive to content changes. It can be applied to audiovisual content authentication in commerce, law, defense, and journalism.

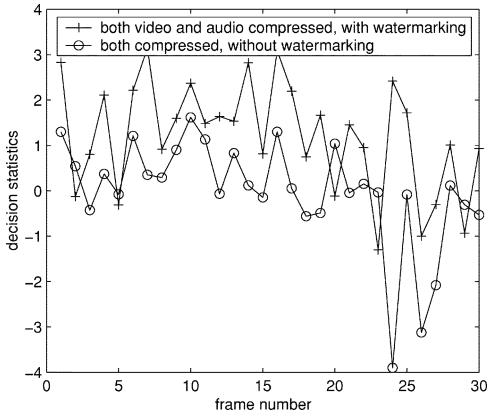


Fig. 19. Decision statistic using Dittmann's method.

For future research, more performance analyses are needed [50], and incorporation of any refined perceptual model for multiplicative watermarks can be investigated.

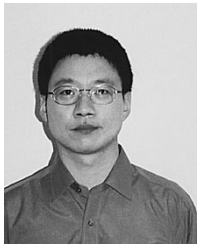
ACKNOWLEDGMENT

The authors would like to thank Prof. P. Moulin, Prof. S. Levinson, and Prof. N. Ahuja at UIUC, Prof. B. Liu at Princeton, and anonymous reviewers for their insightful and invaluable suggestions and comments. The authors would also like to thank Dr. H. Pan at Sharp for discussions on audiovisual applications.

REFERENCES

[1] "Special issue on copyright and privacy protection," *IEEE J. Select. Areas Commun.*, vol. 16, May 1998.
[2] "Special issue on identification and protection of multimedia information," *Proc. IEEE*, vol. 87, July 1999.
[3] I. J. Cox, M. L. Miller, and A. L. McKellips, "Watermarking as communications with side information," *Proc. IEEE*, vol. 87, pp. 1127–1141, July 1999.
[4] M. Ramkumarand and A. N. Akansu, "Capacity estimates for data hiding in compressed images," *IEEE Trans. Image Processing*, vol. 10, pp. 1252–1263, Aug. 2001.
[5] P. Moulin and J. A. O'Sullivan. Information-theoretic analysis of information hiding. preprint, Sept. 1999; revised, Dec. 2001. [Online]. Available: http://www.ifp.uiuc.edu/moulin/.
[6] ——, "Information-theoretic analysis of watermarking," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Istanbul, Turkey, July 2000.
[7] Y. Steinberg and N. Merhav, "Identification in the presence of side information with application to watermarking," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1410–1422, May 2001.
[8] C. Hsu and J. Wu, "Hidden digital watermarks in images," *IEEE Trans. Image Processing*, vol. 8, pp. 58–68, Jan. 1999.
[9] M. Ramkumar and A. N. Akansu, "A robust scheme for oblivious detection of watermarks/date hiding in still images," in *Proc. SPIE Symp. Voice, Video, Data Commun.*, vol. 3528, Boston, MA, Nov. 1998.
[10] M. Yeung and F. Mintzer, "Invisible watermarking for image verification," *J. Electron. Imag.*, vol. 7, no. 3, July 1998.
[11] M. Wu and B. Liu, "Watermarking for image authentication," in *Proc. Int. Conf. Image Process.*, Chicago, IL, Oct. 1998.
[12] M. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 439–441, May 1983.
[13] B. Chen and G. W. Wornell, "An information-theoretic approach to the design of robust digital watermarking systems," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Phoenix, AZ, Mar. 1999.
[14] I. 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.
[15] C. Podilchuk and W. Zeng, "Image-adaptive watermarking using visual models," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 525–539, May 1998.
[16] M. D. Swanson, B. Zhu, and A. H. Tewfik, "Multiresolution scene-based video watermarking using perceptual models," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 540–550, May 1998.
[17] R. B. Wolfgang, C. I. Podilchuk, and E. J. Delp, "Perceptual watermarks for digital images and video," *Proc. IEEE*, vol. 87, pp. 1108–1126, July 1999.
[18] M. Zeng and B. Liu, "A statistical watermark detection technique without using original images for resolving rightful ownerships of digital images," *IEEE Trans. Image Processing*, vol. 6, pp. 1534–1548, Nov. 1999.
[19] G. Langelaar, I. Setyawan, and R. Lagendijk, "Watermarking digital image and video data," *IEEE Signal Processing Mag.*, vol. 17, pp. 20–46, Sept. 2000.
[20] J. R. Hernandez and F. Perez-Gonzalez, "Statistical analysis of watermarking schemes for copyright protection of images," *Proc. IEEE*, vol. 87, pp. 1142–1166, July 1999.
[21] J. Hernandez, M. Amado, and F. Perez-Gonzalez, "DCT-domain watermarking techniques for still images: Detector performance analysis and a new structure," *IEEE Trans. Image Processing*, vol. 9, pp. 55–68, Jan. 2000.
[22] Q. Cheng and T. S. Huang, "Blind digital watermarking for images and videos and performance analysis," in *Proc. Int. Conf. Multimedia Expos.*, New York, Aug. 2000.
[23] ——, "An additive approach to transform-domain information hiding and optimum detection structure," *IEEE Trans. Multimedia*, vol. 3, pp. 273–284, Sept. 2001.
[24] M. Barni, F. Bartolini, A. De Rosa, and A. Piva, "A new decoder for the optimum recovery of nonadditive watermarks," *IEEE Trans. Image Processing*, vol. 10, pp. 755–766, May 2001.
[25] A. Piva, M. Barni, F. Bartolini, V. Cappellini, A. De Rosa, and M. Orlandi, "Improving DFT watermarking robustness through optimum detection and synchronization," in *Proc. ACM Workshop Multimedia Security*, Orlando, FL, Oct. 1999.
[26] J. G. Proakis, *Digital Communication*. New York: McGraw-Hill, 1989.
[27] S. G. Wilson, *Digital Modulation and Coding*. Upper Saddle River, NJ: Prentice-Hall, 1996.
[28] R. J. Clarke, *Transform Coding of Images*. New York: Academic, 1985.
[29] S. Voloshynovskiy, S. Pereira, V. Iquise, and T. Pun, "Attack modeling: Toward a second generation watermarking benchmark," *Signal Process.*, vol. 81, pp. 1177–1214, 2001.
[30] Q. Cheng and T. S. Huang, "Optimum detection of multiplicative watermarks using locally optimum decision rule," in *Proc. Int. Conf. Multimedia Expos.*, Tokyo, Japan, Aug. 2001.
[31] P. Moulin and A. Ivanovic, "The watermarking selection game," in *Proc. Conf. Inform. Sci. Syst.*, Baltimore, MD, Mar. 2001.
[32] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1994.
[33] T. S. Ferguson, *Mathematical Statistics: A Decision Theoretic Approach*. New York: Academic, 1967.
[34] J. H. Miller and J. B. Thomas, "Detectors for discrete-time signals in non-Gaussian noise," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 241–250, 1972.
[35] S. A. Kassam and H. V. Poor, "Robust techniques for signal processing: A survey," *Proc. IEEE*, vol. 73, pp. 433–481, 1985.
[36] M. Hansen and B. Yu, "Wavelet thresholding via MDL for natural images," *IEEE Trans. Inform. Theory*, vol. 46, pp. 1778–1788, Aug. 2000.
[37] H. L. Van Trees, *Detection, Estimation and Modulation Theory*. New York: Wiley, 1968.
[38] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. New York: McGraw-Hill, 1991.
[39] R. W. Buccigrossi and E. P. Simoncelli, "Image compression via joint statistical characterization in the wavelet domain," *IEEE Trans. Image Processing*, vol. 8, pp. 1688–1701, Dec. 1999.
[40] M. Kutter, S. Voloshynovskiy, and A. Herrigel, "The watermark copy attack," *Proc. SPIE Security Watermarking Multimedia Contents II*, vol. 3971, Jan. 2000.
[41] J. Dittmann, T. Fiebig, R. Steinmetz, S. Fischer, and I. Rimac, "Combined video and audio watermarking: Embedding content information in multimedia data," *Proc. SPIE Security Watermarking Multimedia Contents II*, vol. 3971, Jan. 2000.
[42] N. Nocerino, F. K. Soong, L. R. Rabiner, and D. H. Klatt, "Comparative study of several distortion measures for speech recognition," *Speech Commun.*, vol. 4, pp. 317–331, 1985.
[43] L. Rabiner and B.-H. Juang, *Fundamentals of Speech Recognition*. Upper Saddle River, NJ: Prentice-Hall, 1993.

[44] F. Hartung and B. Girod, "Digital watermarking of uncompressed and compressed video," *Signal Process.*, vol. 66, Mar. 1998.
[45] G. Langelaar, R. Lagendijk, and J. Biemond, "Watermarking by DCT coefficient removal: Statistical approach to optimal parameter settings," *Proc. SPIE Security Watermarking Multimedia Contents II*, vol. 3657, Jan. 1999.
[46] R. Dugad and N. Ahuja, "A scheme for joint watermarking and compression of video," in *Proc. Int. Conf. Image Process.*, Sept. 2000.
[47] R. R. Wang, Q. Cheng, and T. S. Huang, "Identifying regions of interest for video watermarking with principle component analysis with multiple cues," in *Proc. ACM Multimedia*, Los Angeles, CA, Oct. 2000.
[48] A. Netravali and B. Haskell, *Digital Pictures*. New York: Plenum, 1988.
[49] F. Deguillaume, G. Csurka, and T. Pun, "Countermeasures for unintentional and intentional video watermarking attacks," in *Proc. SPIE Security Watermarking Multimedia Contents II*, vol. 3971, Jan. 2000.
[50] Q. Cheng and T. S. Huang, "Optimal detection and decoding of multiplicative watermarks in DFT domain," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Orlando, FL, May 2002.



Qiang Cheng (M'03) received the B.S. and M.S. degrees from Peking University, Beijing, China, in 1994 and 1996, respectively. He received the Ph.D. degree from the Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign (UIUC), in 2002.

He is currently an assistant professor with the Electrical and Computer Engineering Department, Wayne State University, Detroit, MI. He held a Graduate School Fellowship at UIUC from 1997 to 1998 and a Guanghua University Fellowship from Peking University from 1995 to 1996. His research interests include multimedia computing, watermarking, communications, information and network security, and statistical learning.

Dr. Cheng received the Antai Award for Academic Excellence from Peking University in 1995 and the Invention Achievement Award from IBM T. J. Watson Research Center, Yorktown Heights, NY, in January 2001.

Thomas S. Huang (LF'01) received the B.S. degree in electrical engineering from National Taiwan University, Taipei, Taiwan, R.O.C. and the M.S. and Sc.D. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge.

He was on the Faculty of the Department of Electrical Engineering at MIT from 1963 to 1973 and on the Faculty of the School of Electrical Engineering and Director of its Laboratory for Information and Signal Processing at Purdue University, West Lafayette, IN, from 1973 to 1980. In 1980, he joined the University of Illinois at Urbana-Champaign, where he is now William L. Everitt Distinguished Professor of Electrical and Computer Engineering, Research Professor at the Coordinated Science Laboratory, and Head of the Image Formation and Processing Group at the Beckman Institute for Advanced Science and Technology and Co-Chair of the Institute's major research theme Human Computer Intelligent Interaction. During his sabbatical leaves, he has worked at the MIT Lincoln Laboratory, Lexington, MA; the IBM Thomas J. Watson Research Center, Yorktown Heights, NY; and the Rheinishes Landes Museum, Bonn, Germany. He has held Visiting Professor positions at the Swiss Institutes of Technology, Zurich and Lausanne; University of Hannover, Hannover, Germany; INRS-Telecommunications of the University of Quebec, Montréal, Canada; and the University of Tokyo, Tokyo, Japan. He has served as a consultant to numerous industrial firms and government agencies both in the U.S. and abroad. His professional interests lie in the broad area of information technology, especially the transmission and processing of multidimensional signals. He has published 12 books and over 400 papers in network theory, digital filtering, image processing, and computer vision. He is a Founding Editor of the *International Journal Computer Vision, Graphics, and Image Processing* and Editor of the Springer Series in Information Sciences, published by Springer Verlag.

Dr. Huang is a Member of the National Academy of Engineering, a Foreign Member of the Chinese Academies of Engineering and Sciences, and a Fellow of the International Association of Pattern Recognition and the Optical Society of America. He has received a Guggenheim Fellowship, an A.V. Humboldt Foundation Senior U.S. Scientist Award, and a Fellowship from the Japan Association for the Promotion of Science. He received the IEEE Signal Processing Society's Technical Achievement Award in 1987 and the Society Award in 1991. He was awarded the IEEE Third Millennium Medal in 2000. In addition, in 2000, he received the Honda Lifetime Achievement Award for "contributions to motion analysis." In 2001, he received the IEEE Jack S. Kilby Medal. In 2002, he received the King-Sun Fu Prize from the International Association of Pattern Recognition and the Pan Wen-Yuan Outstanding Research Award.