Mismatching Perceptual Models for Effective Watermarking in the Presence of Compression

Deepa Kundur and Dimitrios Hatzinakos

Department of Electrical and Computer Engineering
University of Toronto
10 King's College Road
Toronto, Ontario Canada M5S 3G4

ABSTRACT

In this paper, we concentrate on the problem of robust watermarking in the presence of perceptual coding. We first present a watermarking approach called Robust Reference Watermarking (RRW) which can be incorporated into a broad class of watermarking algorithms to improve their performance. Through analysis of this scheme we demonstrate how embedding the watermark using a different perceptual model and domain than for compression can result in improved watermark extraction and detection reliability. Simulation results are also presented to verify our theoretical observations.

Keywords: multimedia security, digital watermarking, perceptual coding, data hiding, steganography.

1. INTRODUCTION

The wide-spread communication of multimedia data has created a growing need to protect digital information against illegal duplication and manipulation. Advances that facilitate electronic publishing and commerce also heighten threats of intellectual property theft and unlawful tampering. One approach to address this problem involves embedding an invisible structure into a *host signal* to "mark" its ownership. We call these structures digital watermarks and the associated embedding process *digital watermarking.* One major driving force for research in this area is the need for effective copyright protection scenarios for digital imagery, sound and video. In such an application a serial number is embedded into the signal to protect to identify the copyright holder. It is expected that an *attacker* will attempt to remove the watermark by intentionally modifying the watermarked signal.

At present, digital watermarking research primarily involves the identification of effective signal processing strategies to discreetly, robustly, and unambiguously hide the watermark information into multimedia signals. The general process involves the use of a key which must be used to successfully embed and extract the hidden information. The embedding mechanism entails imposing imperceptible changes to the host signal to generate a *watermarked signal* containing the watermark information, while the extraction routine attempts to reliably recover the hidden watermark from a possible tampered watermarked signal.

1.1. Previous Work

Initial research into digital watermarking concentrated on the design of sophisticated embedding strategies to improve robustness against typical signal distortions. More recent work has focussed in particular on assessing the effects of perceptual coding on the embedded watermark.[1–4] *Perceptual coding* refers to the *lossy* compression of multimedia signal data using human perceptual models; the compression mechanism is based on the premise that minor modifications of the signal representation will not be noticeable in the displayed signal content. These modifications are imposed on the signal in such a way as to reduce the number of information bits required for storage of the content. Human perceptual models are often theoretically and experimentally derived to determine the changes on a signal which remain imperceptible. A duality exists between the problems of perceptual coding and watermarking; the former problem attempts to remove irrelevant and redundant information from a signal, while the latter uses the irrelevant information to mask the presence of the watermark. Thus, the objectives of the two mechanisms are

Further author information:
D.K.: E-mail: deepa@comm.toronto.edu
D.H.: E-mail: dimitris@comm.toronto.edu

Part of the SPIE Conference on Multimedia Systems and Applications II
Boston, Massachusetts • September 1999 SPIE Vol. 3845 • 0277-786X/99/$10.00

somewhat at odds with one another. As a result, several papers have dealt with integrating perceptual coding with watermarking .[5-10]

There are three basic strategies commonly employed to combine both problems. In the first situation, the watermark is added to the multimedia signal after compression; this normally occurs by imposing changes on the quantized transform coefficients. In this way, the compression algorithm will not effect the reliability of the embedded watermark. The main disadvantage of this approach is that the resulting changes to embed the watermark may result in perceived distortions on the resulting watermarked signal. In fact, one such technique proposed by Meng and Chang[7] embeds a *visible* watermark in the compressed domain instead of attempting to keep the changes invisible.

In the second case, watermarking and perceptual coding are integrated, so that data hiding occurs in the compression domain together with quantization which reduces complexity since only one transform module can be used for both compression and watermarking. Lacy *et al.*[5] proposed such a technique in which they embed the watermark in the compression domain by changing the structure of the associated encoder and by sacrificing some compression efficiency to hide information. The authors suggest that their method is secure against tampering since any modification of the signal coefficients in the compressed domain will cause visual degradation. However, the approach is susceptible to attacks if one possesses a working knowledge of the coder structure.

In the third situation, watermarking is assumed to occur on the raw signal information prior to compression. Both algorithms are assumed to work independently (although the watermarking algorithm can use knowledge of the structure of the coder, if known, to more reliably embed the information), which provides the flexibility to employ different transforms for both watermarking and coding. Podilchuk and Zeng[11] employ models of the human visual system (HVS) used for the perceptual coding stage to embed a robust watermark which is more optimally masked by the host image. Use of HVS models allows the watermark to be adapted to both global viewing conditions as well as local properties of the host signal associated with visual masking. Furthermore, Wolfgang, Podilchuk and Delp[2] suggest that one must use the same transforms for watermarking as for compression to maximize robustness. Their, simulations, however, are inconclusive as the results do not strongly support their hypothesis. Their wavelet-based watermarking scheme is more robust to the JPEG algorithm than their DCT-based watermarking. Furthermore, Inoue *et al.*[1] examine the robustness of a proposed wavelet-based watermarking to JPEG compression and obtain similar results which demonstrates the superiority of watermarking in the wavelet domain for distortions due to JPEG compression.

The general trend in current research is to embed the watermark in the same domain as used for perceptual coding.[3,4,6,9,12] Although it is correct that such a strategy is better in reducing complexity, it is not necessarily true in terms of robustness. The natural question arises as to what type of scheme is fundamentally better when trying to maximize robustness of the watermark for a given coder efficiency.

1.2. Contributions of this Paper

In this paper, we argue both analytically and through the use of simulation results that the use of the same transforms for both watermarking and compression result in performance which is far from optimal. That is, complementary transforms can potentially provide greater robustness. There have been some indications of this in earlier work by the authors[13,14] and in the recent work by Ramkumar and Akansu[15] in which they reason that use of transforms with poor energy compaction properties, which are not suited for compression, provide greater watermark capacity.

This work begins with an introduction of a watermarking approach called Robust Reference Watermarking (RRW) which was initially introduced by the authors of this paper in Ref. 13. The technique exploits diversity principles and channel characterization tools to improve watermark robustness. A communications paradigm for watermarking is established and an effective linear watermark receiver structure to improve watermark recovery is discussed. The basic RRW approach can be applied to a broad class of watermarking algorithms to enhance robustness. We go beyond our previous work[13] and provide rigorous analysis of the RRW algorithm. It is demonstrated that our watermark receiver structure minimizes the probability of bit error. We also derive an error static bound useful in the evaluation of robustness. Section 3 discusses effective watermarking strategies; implications for watermarking in the presence of perceptual coding are specifically addressed. Our theoretical observations are verified in simulations in Sec. 4. Final remarks conclude the paper.

2. ROBUST REFERENCE WATERMARKING

The RRW approach can be applied to many existing robust data hiding algorithms [11,16–20] to improve the reliability of the extracted watermark. In this paper, we assume without loss of generality that the host signal is a still image. Our approach can be incorporated into a broad class of watermarking systems with the following basic characteristics:

1. The watermark w is binary and of length N_w bits which is much smaller than the number of pixels in the host image, which we represent by the vector $\mathbf{f}_0 \in \mathbb{R}^{N_f}$ where N_f is the total number of image pixels. This basic assumption works for most watermarking applications such as copy protection, hidden annotations and covert communications.

2. The watermark information is repeatedly embedded $M \geq 1$ times within the host signal $\mathbf{f}_0$.

3. The embedding process occurs in the *watermark domain*. Specifically, a linear orthogonal one-to-one transformation $T_W : \mathbb{R}^{N_f} \to \mathbb{R}^{N_f}$ is applied to the host signal $\mathbf{f}_0$. The transformation T_W decomposes the host signal vector $\mathbf{f}_0$ into coefficients corresponding to a series of orthonormal N_f-dimensional basis vectors $\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_{N_f}$ to produce the watermark domain coefficient vector $\mathbf{F}_0 = [F_0(1)\ F_0(2)\ \cdots F_0(N_f)]^{\mathrm{T}}$ (where $(\cdot)^{\mathrm{T}}$ is the transpose operator) such that, $\mathbf{f}_0 = \sum_{i=1}^{N_f} F_0(i)\boldsymbol{\phi}_i$. Alternatively, we can write that $F_0(i) =< \mathbf{f}_0, \boldsymbol{\phi}_i >$, where $< \cdot, \cdot >$ is the inner product of the two argument vectors. Many common transformations used in digital signal processing such as the DCT, orthogonal wavelet transforms, the Haar transform, the Hadamard transform, the sine transform, the discrete Fourier transform and the Karhunen Loeve transform fulfill these requirements.

4. The kth repetition of the N_w-bit watermark is embedded (using a routine such as those proposed in Refs. 11,16–20) into the following group of coefficients

$$\mathbf{F}_0^{(k)} = [F_0((k-1)N_s+1)\ F_0((k-1)N_s+2)\ \cdots\ F_0(kN_s)]^{\mathrm{T}} \tag{1}$$

   for $k = 1, 2, \ldots, M$, by modifying them in some imperceptible way to form the corresponding watermarked coefficients

$$\mathbf{F}_W^{(k)} = [F_W((k-1)N_s+1)\ F_W((k-1)N_s+2)\ \cdots\ F_W(kN_s)]^{\mathrm{T}} \tag{2}$$

   for $k = 1, 2, \ldots, M$. The associated watermark domain coefficient vector is denoted $\mathbf{F}_W$.

   Equivalently, each coefficient set of the watermarked signal $\mathbf{F}_W^{(k)}$ contains the complete watermark information. Since M repetitions of the watermark are embedded in the host signal, the coefficients $F_0(1)$, $F_0(2)$, $\ldots$, $F_0(MN_s)$, are modified to embed the watermark information and the remaining coefficients $\{F_0(MN_s+1), F_0(MN_s+2), \ldots, F_0(N_f)\}$ are untouched to form the watermarked signal. We denote the watermarked image vector in the spatial domain by $\mathbf{f}_w$.

5. The watermarked signal may undergo distortions that effect the integrity of the embedded watermark information. If we let $\hat{\mathbf{f}}_w$ represent the possibly modified watermarked signal vector, then to extract the watermark repetitions we form $\hat{\mathbf{F}}_W$ by transforming $\hat{\mathbf{f}}_w$ with T_W. The set of coefficients

$$\hat{\mathbf{F}}_W^{(k)} = [\hat{F}_W((k-1)N_s+1)\ \hat{F}_W((k-1)N_s+2)\ \cdots\ \hat{F}_W(kN_s)]^{\mathrm{T}} \tag{3}$$

   for $k = 1, 2, \ldots, M$, is used to extract the kth watermark repetition. This extracted watermark bit sequence is denoted $\hat{w}_k$ and has an associated reliability factor; we use the probability of bit error measure p_{Ek}.

Any one of a variety of well-known data embedding and extraction routines can be applied to form $\mathbf{F}_W^{(k)}$ from $\mathbf{F}_0^{(k)}$. Moreover, our analysis is independent of such mechanisms. Only during the implementation of our RRW approach described in Sec. 4 do we make use of the embedding method described in Ref. 20 to generate simulation results. Although we restrict the watermark to be a bit sequence and the reliability measure to be the probability of bit error, we believe the spirit of our analytic results holds for non-binary watermarks with a different reliability measure such as the signal-to-noise ratio (SNR).

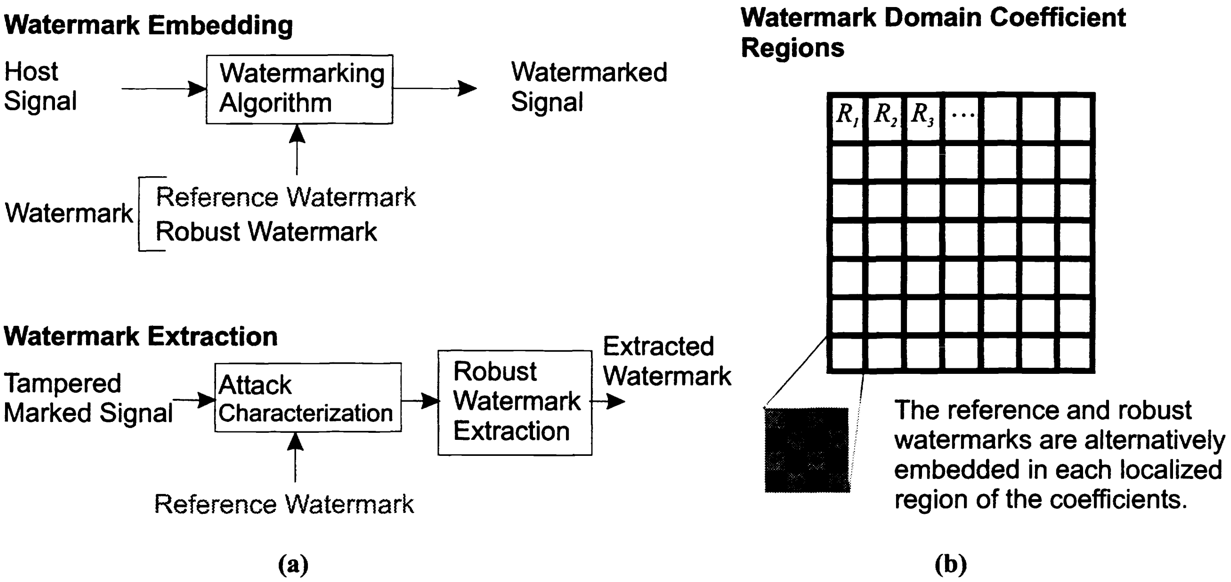

Figure 1. Robust Reference Watermarking. (a) The watermark embedding and extracting scenarios, (b) We consider a 2-D host image. The watermark domain coefficients are divided into localized regions R_k (outlined with bold lines). Reference and robust watermark bits are alternatively embedded in each region.

2.1. The Role of the Reference Watermark

We define a reference watermark as one which is embedded into a signal for the purpose of detecting signal distortions. It has been shown in Ref.[21] that elementary characteristics of signal distortions can be easily estimated using a watermark.

We propose the scenario in which the host signal is embedded with both robust and reference watermarks. The two kinds of watermarks are placed orthogonally so that they do not interfere with one another in the host signal. Specifically, the watermarks are placed in separate host signal coefficients. The trade-off is that fewer repetitions of the N_w-bit robust watermark can be placed in the signal as a portion of the watermark "bandwidth" is consumed by the presence of the reference watermark. Each embedded repetition of the robust watermark sequence, which we denote w_k, $k = 1, 2, \ldots M$ (where M is the total number of repetitions), has an associated N_r-bit binary reference watermark sequence r_k. Since w_k is a repetition of the robust watermark, $w_k = w_j$ for all k and j. The reference watermarks $\{r_k\}$ do not necessarily have to be identical, but for the implementation and simulations in this paper, we let $r_k = r_j$ for all k and j. Figure 1(b) demonstrates the embedding procedure where each w_k and r_k are placed in a localized region denoted R_k of the watermark domain, where k is the associated index. This localized region is a rectangular spatial neighbourhood of transform coefficients. Specifically, in the watermark domain, the coefficients are segmented into non-overlapping blocks to represent localized regions. For each region, R_k, $N_w + N_r$ bits are embedded using an existing robust data hiding algorithm, so that one entire repetition of each of w and r can be embedded in the region. Assuming $N_w = N_r$, the bits of w_k can be alternated with those of r_k in a checker board pattern such that an attack on the marked signal will reflect statistically in the same way on both w_k and r_k. Thus, if we let $\hat{w}_k$ and $\hat{r}_k$ be the extracted versions of w_k and r_k after an attack, we expect that the probability of bit error for $\hat{w}_k$ is equal to that for $\hat{r}_k$.

The approach is similar to the concept of a training sequence or a reference signal used in digital communications in which a known data sequence is transmitted from the source to the destination to characterize the communications channel. Proper identification of our associated *watermark channel* (i.e., the effective channel communicating the watermark through the distortions imposed on the multimediasignal) will allow more accurate extraction of the robust watermark as compensating processing may be incorporated at the receiver. The channel estimation is performed with the use of the reference watermark. The robust watermark is not used for estimation of the watermark channel as it may be unknown at the extraction end and because we intend to obtain an unbiased characterization of the attack. Use of the robust watermark, if it is known, will increase the probability of false positive detection which may not be appropriate for some applications. In the next section, we discuss the particular model of the watermark channel model we assume.

2.2. BSC Model of the Watermark Channel

Since each watermark repetition w_k and its associated reference watermark r_k are embedded in the same localized region R_k of the watermark domain as shown in Fig. 1(b), most degradations which maintain the perceptual quality of the signal will have a similar effect on both w_k and r_k. Therefore, it follows that the degree of distortion experienced by both w_k and r_k due to an attack is the same, hence, they have they same watermark channel.

We model the watermark channel for w_k and r_k as a binary symmetric channel (BSC) with probability of bit error p_{Ek}. Therefore, each bit of the embedded robust watermark $w_k(i)$, $i = 1, 2, \ldots, N_w$, is modeled as passing through a BSC to produce the corresponding extracted watermark bit $\hat{w}_k(i)$. We assume in our model that $0 \leq p_{Ek} \leq 0.5$. If $p_{Ek} > 0.5$ we merely complement the output and effectively use $0 \leq 1 - p_{Ek} < 0.5$ as the BSC parameter.

The reference watermark r_k is used to estimate the parameter p_{Ek} for each k. If we let r_k be the corresponding extracted reference watermark after an attack, we can approximate the probability of bit error for the associated watermark channel by

$$\hat{p}_{Ek} = \frac{1}{N_r} \sum_{i=1}^{N_r} r_k(i) \oplus \hat{r}_k(i) \tag{4}$$

where $\oplus$ is the exclusive-OR operator, and $r_k(i)$ and $\hat{r}_k(i)$ are the ith watermark bits of r_k and $\hat{r}_k$, respectively. It can be shown using the law of large numbers that the expected value of $\hat{p}_{Ek}$ is p_{Ek} and that the variance of estimate decreases for increasing N_r.

There are important advantages to using this model of the watermark channel. The model is simple and the parameter p_{Ek} is easy to accurately estimate using the associated reference watermark. In addition, a different parameter p_{Ek} for each w_k is incorporated which provides a localized assessment of the attack in the wavelet domain. In most watermarking schemes, the extracted watermark repetitions $\hat{w}_k$ are averaged to produce the overall extracted watermark. Our attack characterization allows us to combine these repetitions based on a measure of their reliability to minimize the probability of watermark bit error. Although this characterization is appropriate for degradations such as filtering, additive noise and lossy compression, it should be emphasized that it is not well-suited for geometric transformations on the signal such as rotation and scaling.

2.3. An Effective Watermark Receiver Structure

The distinct watermark repetitions w_k are extracted from the transform coefficients $\hat{\mathbf{F}}_W^{(k)}$ using the appropriate data recovery algorithm. The BSC parameters can be used to obtain a more accurate estimate of the watermark information compared to simple averaging of the extracted repetitions. To keep computational complexity low so we limit ourselves to linear estimation. The overall extracted watermark $\hat{w}$ is computed as the weighted sum of the individual extracted repetitions. That is,

$$\hat{w}(i) = \text{round}\left[\sum_{k=1}^{M} \alpha_k \hat{w}_k(i)\right] \tag{5}$$

where $\hat{w}(i)$ and $\hat{w}_k(i)$ are the ith watermark bits of $\hat{w}$ and $\hat{w}_k$, respectively, and α_k is the associated scalar nonnegative weight dependent on p_{Ek} such that $\sum_{k=1}^{M} \alpha_k = 1$. The rounding operation makes sure that $\hat{w}$ is a binary data string comprised of zeros and ones. If the argument of the round[$\cdot$] is 0.5, an arbitrary value of 0 or 1 is assigned for $\hat{w}(i)$. In any type of watermark attack, some regions in R_k are likely to undergo greater distortion than others. It is a direct advantage to be able to determine the regions which are less distorted (and hence contain a more reliable watermark estimate). It is intuitively clear that a larger weighting for repetitions with a lower probability of bit error will improve the reliability of $\hat{w}$. We show in Appendix A that the following assignment for α_k minimizes the bit error of $\hat{w}$ to produce an optimal linear watermark extraction:

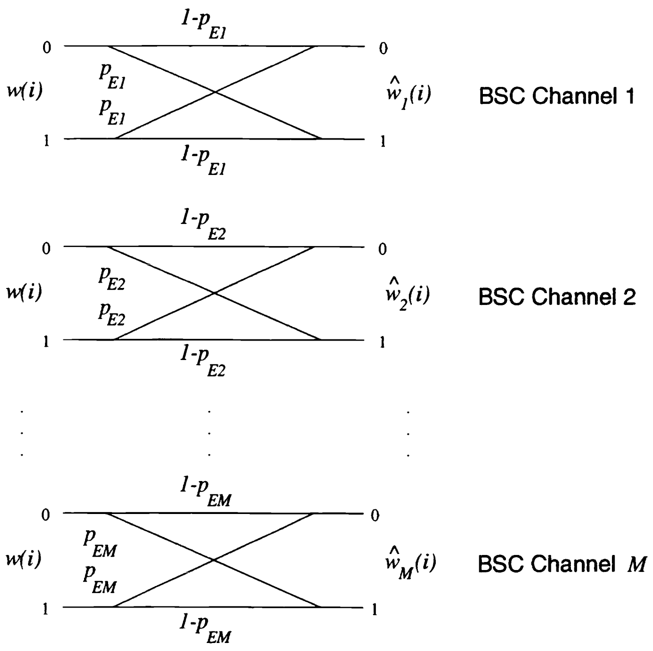

Figure 2. Parallel Binary Symmetric Channel Model.

$$\alpha_k = \frac{\log\left(\frac{1-p_{Ek}}{p_{Ek}}\right)}{\left(\sum_{j=1}^{M}\log\left(\frac{1-p_{Ej}}{p_{Ej}}\right)\right)}. \tag{6}$$

2.4. Analysis

We analyze our proposed watermarking approach to derive fundamental strategies for effective watermarking. Consider the image $\mathbf{f}_w$ from which we intend to recover the embedded watermark. We extract the individual repetitions to produce M estimates of the watermark, $\hat{w}_1, \hat{w}_2, \ldots, \hat{w}_M$.

Our framework is analogous to transmitting the watermark simultaneously along M independent binary symmetric channels as shown in Fig. 2 for which the error probabilities p_{Ek} are assumed to be known; in our technique we estimate p_{Ek} using Eq. (4).

Consider the bit $e_k(i)$ defined as

$$e_k(i) \triangleq w(i) \oplus \hat{w}_k(i) = \begin{cases} 1 & \text{if there is a bit error in } \hat{w}_k(i) \\ 0 & \text{otherwise} \end{cases}. \tag{7}$$

Similarly, we let

$$e(i) \triangleq w(i) \oplus \hat{w}(i) = \begin{cases} 1 & \text{if there is a bit error in } \hat{w}(i) \\ 0 & \text{otherwise} \end{cases}. \tag{8}$$

It follows from Eq. (5) that

$$e(i) = \text{round}\left[\sum_{k=1}^{M} \alpha_k e_k(i)\right], \tag{9}$$

which relates the bit errors of the individual extracted repetitions to the bit error of the overall watermark estimate. Analysis of Eq. (9) is not straightforward due to the presence of the integer round operator. Alternatively, we consider the argument of this operator which is given by

$$\mathcal{E}(\mathbf{e}(i)) \triangleq \sum_{k=1}^{M} \alpha_k e_k(i) \tag{10}$$

where $\mathbf{e}(i) = [e_1(i)\ e_2(i)\ \cdots\ e_M(i)]^{\mathrm{T}}$. A bit error occurs in $\hat{w}(i)$ if $\mathcal{E}(\mathbf{e}(i)) > 0.5$. We can analyze the mean value of $\mathcal{E}(\mathbf{e}(i))$, $E\{\mathcal{E}(\mathbf{e}(i))\}$, to assess the reliability of the watermark channel. Although this is not a precise measure of the error rate of the system since a smaller $E\{\mathcal{E}(\mathbf{e}(i))\}$ does not necessarily guarantee a lower overall bit error rate, it does provide some useful insight into the watermarking problem.

It is shown in Appendix B that

$$E\{\mathcal{E}(\mathbf{e}(i))\} \leq \frac{\overline{p}_E}{1-\overline{p}_E}\left[1 - \frac{D(q_a||q_b)}{\log\left(\frac{1-\overline{p}_E}{\overline{p}_E}\right)}\right] \tag{11}$$

where

$$\overline{p}_E = \frac{1}{M}\sum_{k=1}^{M} p_{Ek}, \tag{12}$$

$D(q_a||q_b)$ is the relative entropy given by[22]

$$D(q_a||q_b) = \sum_{k=1}^{M} q_a(k)\log\left(\frac{q_a(k)}{q_b(k)}\right), \tag{13}$$

$$q_a(k) = p_{Ek}/(M\overline{p}_E), \tag{14}$$

and

$$q_b(k) = (1-p_{Ek})/(M(1-\overline{p}_E)). \tag{15}$$

We can see that q_a and q_b are probability-like distributions since their elements are nonnegative and sum to one. The bound of (11) is tight for small $\overline{p}_E$ and p_{Ek} close to a constant. Specifically, the equality of (11) holds if and only if $p_{Ek} = 0$ for all k.

3. IMPLICATIONS FOR WATERMARKING IN THE PRESENCE OF PERCEPTUAL CODING

3.1. Effective Watermarking Strategies for Arbitrary Distortions

Analysis of (11) reveals that the following possible tactics may be incorporated into a watermarking scheme to lower the value of $E\{\mathcal{E}(\mathbf{e}(i))\}$ and, hence, improve the robustness of the watermarking system in some way:

Reduce the value of $\overline{p}_E$. Reducing the value of $\overline{p}_E$ decreases the term $\frac{\overline{p}_E}{1-\overline{p}_E}$ and increases the denominator term $\log\left(\frac{1-\overline{p}_E}{\overline{p}_E}\right)$ which both serve to lower the overall bound. Many currently proposed watermarking methods attempt to gain performance by diminishing this average bit error rate. Signal processing strategies to imperceptibly embed a higher energy and, hence, more robust watermark are commonly employed.

Embed the watermark such that the distributions q_a and q_b are dissimilar for a large class of distortions. Given a fixed value of $\overline{p}_E$, we may reduce the performance bound by increasing the value of $D(q_a||q_b)$. The relative entropy $D(\cdot||\cdot)$ is a measure of the distance between its two argument distributions.[22] Roughly, we can see that $D(q_a||q_b)$ is large when $q_a(k) = p_{Ek}/(M\overline{p}_E)$ and $q_b(k) = (1 - p_{Ek})/(M(1 - \overline{p}_E))$ are dissimilar. Assuming a fixed $\overline{p}_E$, this requires that p_{Ek} vary in amplitude for different values of k, implying that we should embed the watermark in a domain for which the degree of distortion varies in each localized region corresponding to the coefficients $\mathbf{F}_0^{(k)}$. This can be achieved by embedding the watermark in a domain which distributes the distortion more to certain coefficients, leaving others less effected.

Localize the distortions on the watermarked signal. It is intuitively clear from the examination of Eqs. (5) and (6) that the existence of $p_{Ek} = 0$ for at least one $k \in \{1, 2, \ldots, M\}$ implies that $\mathcal{E}(\mathbf{e}(i)) = 0$. Thus, if there exists a set of coefficients $\hat{\mathbf{F}}_W^{(k)}$ for some k which are unmodified by the distortion, then perfect watermark recovery is possible, even if $\overline{p}_E \neq 0$ as long as all $\{p_{Ek}\}$ are known. This translates to embedding the watermark in a domain which completely localizes the distortion to a few coefficients.

3.2. Specific Implications for Perceptual Watermarking and Coding

The main motivation to incorporate models of the HVS in watermarking algorithms is to embed a higher energy, yet perceptually masked signal. Maximization of the watermark amplitude has the effect of minimizing the overall SNR of the extracted watermark. This translates to minimization of the average probability of bit error $\overline{p}_E$. Thus, many proposed algorithms conveniently borrow well-established transforms and HVS models used for perceptual coding for watermarking. This is a practical advantage because it reduces the overall complexity of combining the tasks. However, some compression efficiency must be sacrificed to embed the watermark information reliably.

Given our discussion in Sec. 3.1, it follows that further improvements in watermark robustness can be achieved by embedding the watermark such that the distributions $q_a(k) = p_{Ek}/(M\overline{p}_E)$ and $q_b(k) = (1 - p_{Ek})/(M(1 - \overline{p}_E))$ are dissimilar. For watermark embedding, the perceptual model is employed to determine the maximum strength of the signal which can be imperceptibly embedded in the host signal. For compression, the model determines the maximum amplitude of quantization noise that a coefficient of the signal will experience. Use of the same model for both tasks will, therefore, hold the watermark signal energy proportional to the quantization noise level. It follows that each repetition of the watermark will experience the same probability of bit error and $p_{Ek} = \overline{p}_E$ which implies that $q_a(k) = q_b(k) = 1/M$, so that the error statistic bound of (11) is maximized for a given $\overline{p}_E$. Thus, it appears that use of the same transforms and perceptual models for coding and compression is far from optimal.

This new insight stems from our use of a localized degradation model for the overall watermark channel. Through the use of a parallel BSC structure with different probability of bit error rate for each sub-channel (which is estimated with the use of the reference watermark), we are able to gain more insight into the use of different transforms. This is in contrast to previous work which has assumed that the effects of attacks on the watermark is equivalent to uniform additive white Gaussian noise. Such a model is not instructive to use of different domains for more robust watermarking. The only apparent way to improve performance is to increase the watermark signal energy.

Based on our analysis in this paper, we allege that *complementary* transforms and perceptual models for watermarking and coding will result in superior performance. We do not discuss appropriate design strategies in this paper, but we suggest that one can develop compatible perceptual models for watermarking and coding which incorporate distinct masking characteristics, and hence, are somewhat complementary. Another simple approach to improve performance of watermarking in the presence of coding is to embed the watermark in a *different* domain than used for coding. This is in direct conflict with existing principles and conventions[2,5,7,11] used for effective watermarking. We verify this hypothesis with simulation results presented in the next section. We demonstrate how selection of arbitrarily mismatched transforms can provide better results than equating both domains.

4. SIMULATIONS

We apply our proposed RRW approach to the wavelet-based data hiding technique proposed by the authors of this paper in Ref. 20. In the simulations using the algorithm of Ref. 20, we specifically make use of the Daubechies 10-point wavelet[23] (symbolized with **db5** in Matlab) which is an intentionally arbitrary selection. Parameter values of $L = 4$ and $Q = 3$ were employed which provide a good trade-off between imperceptibility and robustness.

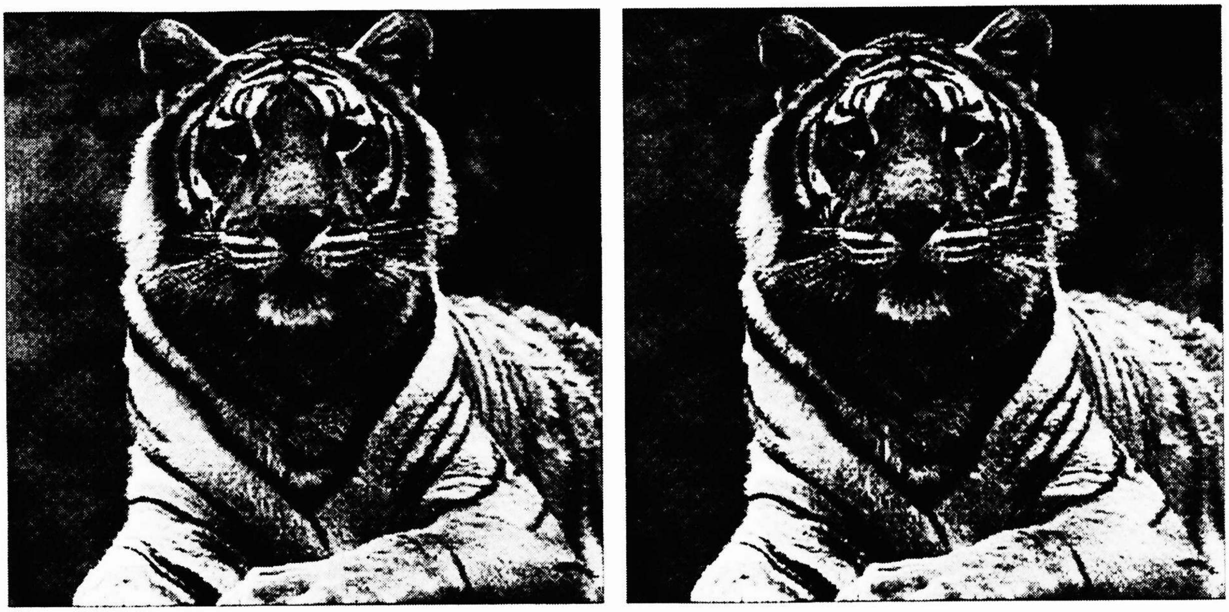

Figure 3. Left: The original host tiger image. **Right:** The watermarked image using the proposed RRW method.

Figure 3 displays the 256 x 256 pixel host and watermarked images used in our simulations. A 128 bit watermark, randomly generated with a uniform probability distribution, was embedded in the host image using the proposed RRW technique. Further details of the implementation are detailed in Ref. 24.

The watermarked image was compressed using a number of different wavelet transforms. Both, the watermarking algorithm and the compression routine were implemented using Matlab. Thirty-seven different wavelet transforms with global hard coefficient thresholding were employed. Specifically, lossy compression involved setting to zero all wavelet coefficient values below a specified threshold $\mathcal{T}$. Six different thresholds were used for each wavelet transform to observe the effects of lossy compression of varying degrees on the embedded watermarking. The first threshold was wavelet dependent and was experimentally determined by the authors to cause negligible perceptual distortion during compression; it's value ranges between 8 and 13. Other constant threshold values of 10, 15, 20, 25 and 30 were tested as well. Global thresholding was used opposed to quantization tables to simplify the comparison process among the different wavelet transforms. Different quantization tables for each transform are difficult to find and their presence would add an additional level of variability to the results which would make assessing the potential of mismatching transforms difficult.

After the watermarked image was compressed using each transform and threshold value, the watermark was extracted and the Hamming distance between the embedded and extracted watermarks was calculated. The Hamming distance measure is given by

$$\rho_{HD}(w, \hat{w}) = \sum_{i=1}^{N_w} w(i)\hat{w}(i) \tag{16}$$

where w and $\hat{w}$ are the embedded and extracted watermarks, respectively. Table 1 displays the results. The left most column shows the particular wavelet type as specified in Matlab notation. The symbol db$N$ represents the Daubechies $2N$-point wavelet, sym$N$ is the $2N$ length wavelet with least asymmetry and the highest number of vanishing moments, coif$N$ represents the $6N$ length Coiflet wavelets, and bior$N_r.N_d$ is the Biorthogonal wavelet with reconstruction order N_r and decomposition order N_d. More information about the wavelets and their implementation, as well as the wavelet compression algorithm can be found in Ref. 25. The second column from the left lists the experimentally determined values for $\mathcal{T}_{OPT}$. The remaining columns contain the Hamming distance values between the embedded and extracted watermarks. A larger value of the Hamming distance implies that the extracted watermark has a greater number of bit errors. The boldface row, corresponding to db5 are the results for compression with the same transform as used for watermarking. As we can see, the results support our hypothesis that use of

Table 1. The Hamming distances between the embedded and extracted watermarks for different wavelets transforms and threshold values used for lossy compression. The optimal threshold value $\mathcal{T}_{OPT}$ is the maximum value of the global threshold for which the degradation on the watermarked image is negligible.

Wavelet	$\mathcal{T}_{OPT}$	$\mathcal{T} = \mathcal{T}_{OPT}$	$\mathcal{T} = 10$	$\mathcal{T} = 15$	$\mathcal{T} = 20$	$\mathcal{T} = 25$	$\mathcal{T} = 30$
db1	11	1	0	3	8	22	23
db2	13	2	0	4	3	17	23
db3	12	3	1	4	10	29	30
db4	12	2	0	1	9	11	14
db5	**12**	**3**	**0**	**2**	**7**	**11**	**25**
db6	12	1	0	3	6	13	14
db7	11	0	1	3	9	21	25
db8	10	1	1	1	8	9	22
db9	11	0	2	6	5	22	21
db10	12	2	2	0	9	15	27
sym2	12	1	0	4	3	17	23
sym3	12	3	1	4	10	29	30
sym4	11	4	2	6	11	21	28
sym5	12	1	0	3	13	21	25
sym6	11	2	2	3	9	11	16
sym7	12	2	2	3	10	8	27
sym8	12	1	0	1	7	9	10
coif1	12	2	2	5	9	15	26
coif2	12	0	1	3	5	10	12
coif3	11	0	1	3	13	15	30
coif4	12	0	0	5	5	7	17
coif5	12	1	1	3	9	18	25
bior1.1	13	2	0	3	8	22	23
bior1.3	13	1	0	1	6	9	14
bior1.5	12	0	0	4	2	3	7
bior2.2	13	3	1	4	11	16	22
bior2.4	13	2	1	1	10	10	16
bior2.6	13	3	0	0	3	7	8
bior2.8	13	2	1	1	1	5	8
bior3.1	8	1	5	13	27	33	46
bior3.3	9	1	0	1	5	12	9
bior3.5	9	2	1	0	9	10	17
bior3.7	10	0	0	2	1	4	9
bior3.9	13	2	1	2	4	9	12
bior4.4	12	1	2	2	10	17	32
bior5.5	12	2	1	7	20	29	46
bior6.8	12	2	1	2	5	10	19

the same transforms can lead to sub-optimal results. Some of the most accurate watermark extractions occur using specific Biorthogonal filters for compression.

Thus, mismatching the transforms, if appropriately selected, can improve the robustness of the watermark in the presence of perceptual coding. The compression efficiency does not necessarily have to be sacrificed to embed the watermark information. Future work involves analytic research to determine appropriate domain pairs.

5. CONCLUSIONS

In this paper we presented the Robust Reference Watermarking approach and introduced analysis to determine effective watermarking strategies in the presence of perceptual coding. The watermarking problem was paralleled to transmitting the watermark bit stream over a series of parallel BSCs. We demonstrate how it is not only necessary to embed a higher energy watermark to improve robustness, but it is also better to do it in a domain which distributes the distortions from perceptual coding more to certain coefficients, leaving others less effected; ideally, we want the distortions contained to a relatively small fraction of the coefficients. By assuming a localized non-uniform degradation model for the watermark we gained new insight into the effect of specific domains on the embedded watermark. The overall success of our proposed RRW approach stems from the use of a reference watermark which locally characterizes distortions on the watermarked signal. We reason that watermarking may be made more robust to perceptual coding by employing complementary perceptual models for coding and watermarking. Simulation results validate our theoretical observations.

APPENDIX A. DERIVATION OF OPTIMAL WEIGHTS

In this section we show how Eq. (5) for α_k given by Eq. (6) minimizes the probability of bit error. We define the following: $e_k(i) \triangleq w(i) \oplus \hat{w}_k(i)$, and $e(i) \triangleq w(i) \oplus \hat{w}(i)$. If $e_k(i) = 1$, there is a bit error in the ith bit of the kth extracted watermark repetition. Eq. (5) becomes

$$e(i) = \text{round}\left[\sum_{k=1}^{M} \alpha_k e_k(i)\right]. \tag{17}$$

We introduce the following *error statistic function*

$$\mathcal{E}(\mathbf{e}(i)) \triangleq \sum_{k=1}^{M} \alpha_k e_k(i). \tag{18}$$

A bit error in $\hat{w}(i)$ occurs if $\mathcal{E}(\mathbf{e}(i)) > 0.5$. For convenience we drop the index i for the remainder of the analysis. We define the following bit error vector

$$\mathbf{e} = [e_1 \; e_2 \; \cdots \; e_M]^T. \tag{19}$$

Because we assume a BSC model for w_k, it follows that e_k is a Bernoulli random variable with probability of "success" (i.e., $P\{e_k = 1\}$) of p_{Ek}. Therefore,

$$P\{\mathbf{e}\} = \prod_{k=1}^{M} {p_{Ek}}^{e_k}(1 - p_{Ek})^{\bar{e}_k}. \tag{20}$$

Our BSCs are assumed to be independent of one another so that e_k and e_j are independent random variables for $k \neq j$.

Theorem 1. Given $\sum_{k=1}^{M} \alpha_k = 1$,

$$\mathcal{E}(\mathbf{e}) = 1 - \mathcal{E}(\bar{\mathbf{e}}) \tag{21}$$

Furthermore,

$$\mathcal{E}(\mathbf{e}) < 0.5 \text{ implies that } \mathcal{E}(\bar{\mathbf{e}}) > 0.5 \tag{22}$$

where $\mathbf{e} = [e_1 \; e_2 \; \cdots \; e_M]^T$ and $\bar{\mathbf{e}} = [\bar{e}_1 \; \bar{e}_2 \; \cdots \; \bar{e}_M]^T$.

Proof

$$\begin{aligned}
1 &= \sum_{k=1}^{M} \alpha_k \qquad (23)\\
&= \sum_{k=1}^{M} \alpha_k (e_k + \overline{e}_k) \qquad (24)\\
&= \sum_{k=1}^{M} \alpha_k e_k + \sum_{k=1}^{M} \alpha_k \overline{e}_k \qquad (25)\\
&= \mathcal{E}(\mathbf{e}) + \mathcal{E}(\overline{\mathbf{e}}) \qquad (26)\\
& \qquad (27)
\end{aligned}$$

Therefore,

$$\mathcal{E}(\mathbf{e}) = 1 - \mathcal{E}(\overline{\mathbf{e}}). \tag{28}$$

Furthermore, If $\mathcal{E}(\mathbf{e}) < 0.5$, then $1 - \mathcal{E}(\overline{\mathbf{e}}) < 0.5$ which implies that $\mathcal{E}(\overline{\mathbf{e}}) > 0.5$. □

Theorem 2.

$$\mathcal{E}(\mathbf{e}) < 0.5 \Longrightarrow P\{\mathbf{e}\} > P\{\overline{\mathbf{e}}\} \tag{29}$$

for

$$\alpha_k = \frac{\log\left(\frac{1-p_{Ek}}{p_{Ek}}\right)}{\left(\sum_{j=1}^{M} \log\left(\frac{1-p_{Ej}}{p_{Ej}}\right)\right)}. \tag{30}$$

Proof

From Theorem 1, $\mathcal{E}(\mathbf{e}) < 0.5$ implies that $\mathcal{E}(\mathbf{e}) < \mathcal{E}(\overline{\mathbf{e}})$. We can, therefore, write

$$\begin{aligned}
& \mathcal{E}(\mathbf{e}) < \mathcal{E}(\overline{\mathbf{e}}) \qquad (31)\\
\Longleftrightarrow\ & \sum_{k=1}^{M} \alpha_k e_k < \sum_{k=1}^{M} \alpha_k \overline{e}_k \qquad (32)\\
\Longleftrightarrow\ & \sum_{k=1}^{M} \log\left(\frac{1-p_{Ek}}{p_{Ek}}\right) e_k < \sum_{k=1}^{M} \log\left(\frac{1-p_{Ek}}{p_{Ek}}\right) \overline{e}_k \qquad (33)\\
\Longleftrightarrow\ & \prod_{k=1}^{M} \left(\frac{1-p_{Ek}}{p_{Ek}}\right)^{e_k} < \prod_{k=1}^{M} \left(\frac{1-p_{Ek}}{p_{Ek}}\right)^{\overline{e}_k} \qquad (34)\\
\Longleftrightarrow\ & \prod_{k=1}^{M} {p_{Ek}}^{\overline{e}_k} (1-p_{Ek})^{e_k} < \prod_{k=1}^{M} {p_{Ek}}^{e_k} (1-p_{Ek})^{\overline{e}_k} \qquad (35)\\
\Longleftrightarrow\ & P\{\overline{\mathbf{e}}\} < P\{\mathbf{e}\} \qquad (36)
\end{aligned}$$

□

From Theorem 1 we can see that either $\mathbf{e}$ or $\overline{\mathbf{e}}$ will result in a bit error for the extracted watermark. Thus, the only way to minimize the probability of bit error would be to set α_k to ensure that the sequence in the pair $\mathbf{e}$ and $\overline{\mathbf{e}}$

with the higher probability will result in no bit error or, equivalently, will result in $\mathcal{E} < 0.5$. Theorem 2 demonstrates that Eq. (6) (also shown in Eq. (30)) achieves this task.

We can form the following set A:

$$A = \{\mathbf{e} \in A \ \mid \ \mathcal{E}(\mathbf{e}) > 0.5\}\,. \tag{37}$$

If the extracted watermark has an associated vector $\mathbf{e}$ such that $\mathbf{e} \in A$ then a bit error will result in the extracted watermark.

APPENDIX B. DERIVATION OF ERROR STATISTIC BOUND

We prove (11) in this section assuming $p_{Ek} \leq 0.5$ for $k = 1, 2, \ldots, M$. From the independence of e_k.

$$\begin{aligned} E\{\mathcal{E}(\mathbf{e})\} &= \frac{\sum_{k=1}^{M} \log\left(\frac{1-p_{Ek}}{p_{Ek}}\right) E\{e_k\}}{\sum_{k=1}^{M} \log\left(\frac{1-p_{Ek}}{p_{Ek}}\right)} && (38) \\ &= \frac{\sum_{k=1}^{M} p_{Ek} \log\left(\frac{1-p_{Ek}}{p_{Ek}}\right)}{\sum_{k=1}^{M} \log\left(\frac{1-p_{Ek}}{p_{Ek}}\right)} && (39) \end{aligned}$$

Using the facts that $\log\left(\frac{1-p_{Ek}}{p_{Ek}}\right) \geq (1 - p_{Ek}) \log\left(\frac{1-p_{Ek}}{p_{Ek}}\right)$ for $p_{Ek} \leq 0.5$

$$E\{\mathcal{E}(\mathbf{e})\} \leq \frac{\sum_{k=1}^{M} p_{Ek} \log\left(\frac{1-p_{Ek}}{p_{Ek}}\right)}{\sum_{k=1}^{M} (1 - p_{Ek}) \log\left(\frac{1-p_{Ek}}{p_{Ek}}\right)} \tag{40}$$

with equality if and only if and only if $p_{Ek} = 0$ for all k. Using the log-sum inequality[22] to the denominator, we can show

$$E\{\mathcal{E}(\mathbf{e})\} \leq \frac{\sum_{k=1}^{M} p_{Ek} \log\left(\frac{1-p_{Ek}}{p_{Ek}}\right)}{M(1 - \overline{p}_E) \log\left(\frac{1-p_{Ek}}{p_{Ek}}\right)} \tag{41}$$

where $\overline{p}_E = \frac{1}{M} \sum_{k=1}^{M} p_{Ek}$ with equality if and only if $p_{Ek} = 0$ for all k.

The right hand side of (41) can be expanded, rearranged, factored and reduced to give (11) which we present again below.

$$E\{\mathcal{E}(\mathbf{e}(i))\} \leq \frac{\overline{p}_E}{1 - \overline{p}_E} \left[1 - \frac{D(q_a \| q_b)}{\log\left(\frac{1-\overline{p}_E}{\overline{p}_E}\right)}\right]$$

where $D(q_a \| q_b)$ is the relative entropy measure given in Eq. (13), and $q_a(k) = p_{Ek}/(M\overline{p}_E)$ and $q_b(k) = (1 - p_{Ek})/(M(1 - \overline{p}_E))$.

ACKNOWLEDGMENTS

The authors gratefully acknowledge Communications and Information Technology Ontario (CITO) for providing financial support for this work. The first author would also like to thank the National Sciences and Engineering Research Council (NSERC) of Canada for providing partial funding for this research in the form of a postgraduate scholarship.

REFERENCES

1. H. Inoue, A. Miyazaki, A. Yamamoto, and T. Katsura, "A digital watermark based on the wavelet transform and its robustness on image compression," in *Proc. IEEE Int. Conference in Image Processing*, vol. 2, October 1998.
2. R. B. Wolfgang, C. I. Podilchuk, and E. J. Delp, "The effect of matching watermark and compression transforms in compressed color images," in *Proc. IEEE Int. Conference in Image Processing*, vol. 1, October 1998.
3. V. Darmstaedter, J.-F. Delaigle, J. J. Quisquater, and B. Macq, "Low cost spatial watermarking," *Computers & Graphics* **22**(4), pp. 417–424, 1998.
4. A. G. Bors and I. Pitas, "Image watermarking using block site selection and DCT domain constraints," *Optics Express* **3**, pp. 512–523, December 7 1998.
5. J. Lacy, S. R. Quackenbush, A. R. Reibman, D. Shur, and J. H. Snyder, "On combining watermarking with perceptual coding," in *Proc. Int. Conf. on Acoustics, Speech and Signal Processing*, vol. 6, pp. 3725–3728, May 1998.
6. H.-J. Wang and C.-C. J. Kuo, "An integrated progressive image coding and watermark system," in *Proc. Int. Conf. on Acoustics, Speech and Signal Processing*, vol. 6, pp. 3721–3724, March 1998.
7. J. Meng and S.-F. Chang, "Embedding visible video watermarks in the compressed domain," in *Proc. IEEE Int. Conference in Image Processing*, vol. 1, 1998.
8. S. Bhattacharjee and M. Kutter, "Compression tolerant image authentication," in *Proc. IEEE Int. Conference in Image Processing*, vol. 1, October 1998.
9. T.-Y. Chung, M.-S. Hong, Y.-N. Oh, D.-H. Shin, and S.-H. Park, "Digital watermarking for copyright protection of MPEG2 compressed video," *IEEE Transactions on Consumer Electronics* **44**, pp. 895–901, August 1998.
10. J. Lacy, S. R. Quackenbush, A. R. Reibman, and J. H. Snyder, "Intellectual property protection systems and digital watermarking," *Optics Express* **3**, pp. 478–484, December 7 1998.
11. C. I. Podilchuk and W. Zeng, "Image-adaptive watermarking using visual models," *IEEE Journal on Selected Areas in Communications* **16**, pp. 525–539, May 1998.
12. C.-T. Hsu and J.-L. Wu, "DCT-based watermarking for video," *IEEE Transactions on Consumer Electronics* **44**, pp. 206-216, February 1998.
13. D. Kundur and D. Hatzinakos, "Improved robust watermarking through attack characterization," *Optics Express* **3**, pp. 485–490, December 7 1998.
14. D. Kundur and D. Hatzinakos, "Attack characterization for effective watermarking," to appear in *Proc. IEEE Int. Conference in Image Processing*, October 1999.
15. M. Ramkumar and A. N. Akansu, "Theoretical capacity measures for data hiding in compressed images," in *Proc. SPIE, Voice, Video and Data Communications*, November 1998.
16. I. J. Cox, J. Killian, T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," Tech. Rep. 95–10, NEC Research Institute, 1995.
17. E. Koch and J. Zhao, "Towards robust and hidden image copyright labeling," in *Proc. Workshop on Nonlinear Signal and Image Processing*, I. Pitas, ed., pp. 452–455, June 1995.
18. J. Ohnishi and K. Matsui, "Embedding a seal into a picture under orthogonal wavelet transform," in *Proc. Int. Conference on Multimedia Computing and Systems*, pp. 514–521, June 1996.
19. G. W. Braudaway, "Protecting publicly-available images with an invisible image watermark," in *Proc. IEEE Int. Conference on Image Processing*, vol. 1, pp. 524–527, October 1997.
20. D. Kundur and D. Hatzinakos, "Digital watermarking using multiresolution wavelet decomposition," in *Proc. IEEE Int. Conference on Acoustics, Speech and Signal Processing*, vol. 5, pp. 2969–2972, 1998.
21. D. Kundur and D. Hatzinakos, "Digital watermarking for telltale tamper-proofing and authentication," in *Proceedings of the IEEE - Special Issue on Identification and Protection of Multimedia Data*, pp. 1167–1180, July 1999.
22. T. Cover and J. Thomas, *Elements of Information Theory*, John Wiley & Sons, Inc., Toronto, 1991.
23. I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Communications in Pure and Applied Mathematics* **41**, pp. 909–996, November 1988.
24. D. Kundur, *Multiresolution Digital Watermarking: Algorithms and Implications for Multimedia Signals.* PhD thesis, University of Toronto, Toronto, Canada, August 1999.
25. M. Misiti, Y. Misiti, G. Oppenheim, and J.-M. Poggi, *Wavelet Toolbox User's Guide*, The MathWorks, Inc., Massachusetts, 1996.