

Watermarking with Diversity: Insights and Implications

Deepa Kundur
University of Toronto

Using communication theory tool sets based on diversity and channel estimation improves the performance of robust digital watermarking algorithms. This article demonstrates that algorithms employing some form of watermark redundancy can be significantly enhanced. It also discusses effective data-hiding strategies within the proposed framework and reports on the results of robust reference watermarking.

Digital watermarking is a useful tool for multimedia security applications such as tamper proofing and assessment, copy control, and fingerprinting. In essence, we can imperceptibly embed a low-energy signal, called a watermark, containing information such as a security code or useful public tags in a host multimedia signal to enhance value. The task of passing a watermark discreetly through a signal is somewhat analogous to the problem of hiding and later finding an almost invisible needle in a rather hefty haystack. A more sophisticated analogy involves communication theory, which researchers have used to design, analyze, and develop performance bounds on robust digital watermarking algorithms. Figure 1 demonstrates the analogy between watermarking and communications. The process of watermark embedding is analogous to channel coding, where the watermark channel is characterized by the distortions on the watermarked signal such as compression or filtering, and watermark detection serves the role of a communication receiver.

Much of the initial work on robust watermarking is based on spread-spectrum principles.¹⁻³ In spread-spectrum watermarking, the embedded signal is a low-energy, pseudorandomly generated white-noise sequence. We can detect it by correlating the known watermark sequence with either an extracted watermark or a transformed version of the watermarked signal. If the correlation fac-

tor is above a given threshold, then we've detected the watermark. The antijamming properties of spread-spectrum signaling make it attractive for watermarking because we can embed a low-energy (and hence imperceptible) watermark—robust to narrow-band interference.

However, spread-spectrum approaches have the following limitations. They allow detection of a known watermark, but the large bandwidth requirement doesn't facilitate extracting a long bit sequence or logo from an audio signal or an image.

Spread-spectrum approaches are specifically vulnerable to the near-far problem.⁴ For watermarking, this implies that if the watermark's energy is reduced because of fading distortions, any residual correlation between the host and watermark can result in unreliable detection.

In addition, many spread-spectrum approaches aren't adaptive. They often don't consider spatial nonstationarity of the host signal and attack interference nor readily incorporate adaptive techniques to estimate the statistical variations.

Furthermore, the correlator receiver structures used for watermark detection aren't effective when fading is present. Although spread-spectrum systems in general try to exploit spreading to average the fading, the techniques aren't designed to maximize performance.

In this work, my colleagues and I hypothesize that some common multimedia signal distortions including cropping, filtering, and perceptual coding aren't accurately modeled as narrow-band interference.⁵⁻⁷ Instead, we assert that such signal modifications cause fading on the embedded watermark. As a result, we can make the watermark more robust by employing effective diversity techniques and channel estimation. We propose a framework that applies to data embedding in most general multimedia signals.

Our analysis is rooted in a family of methods incorporating watermark redundancy. The goal is to develop ideas for enhanced watermarking that exhibit a good compromise among practicality, portability, and general insight. Note that the concepts here are meant to be employed within existing watermarking techniques and aren't intended to replace well-established watermarking strategies such as spread-spectrum watermarking and modulation.

Principles and paradigms

This section explains some of the key concepts related to the foundations of our work.

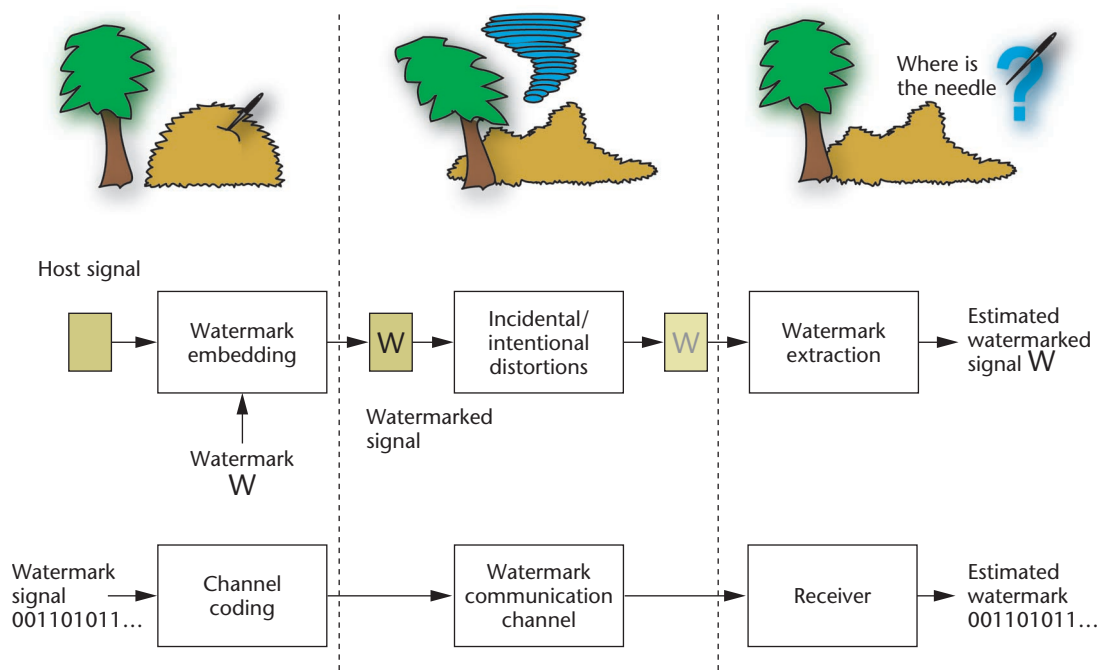


Figure 1. Watermarking analogies. We can model the watermark embedding problem as channel coding in a communication system. The attacks and distortions on the watermark extraction serve the role of the receiver.

Fading

Previous analytic work in the robust watermarking area has assumed stationary additive Gaussian watermark channels. That is, the effect (on the embedded watermark) of distortions on the overall watermarked signal is modeled as additive Gaussian noise. However, some degradation such as cropping or heavy linear filtering completely destroy the watermark content in the signal's associated components. For example, cropping the image to half of its original size or lowpass filtering will annihilate those watermark signal components in the signal's discarded area.

Although certain degradations are more appropriately modeled as fading, we don't assume a particular model for them. However, we believe that the traditional characteristics of a general fading process such as nonstationarity and the need for channel characterization apply. It follows then that basic strategies and rules of thumb to overcome fading in communication theory might also provide performance improvements for robust watermarking.

Diversity

One general way to improve reliability in an unknown, nonstationary environment is to employ diversity. This approach involves transmitting the same information through multiple subchannels of a hostile communications environment to better guarantee information recep-

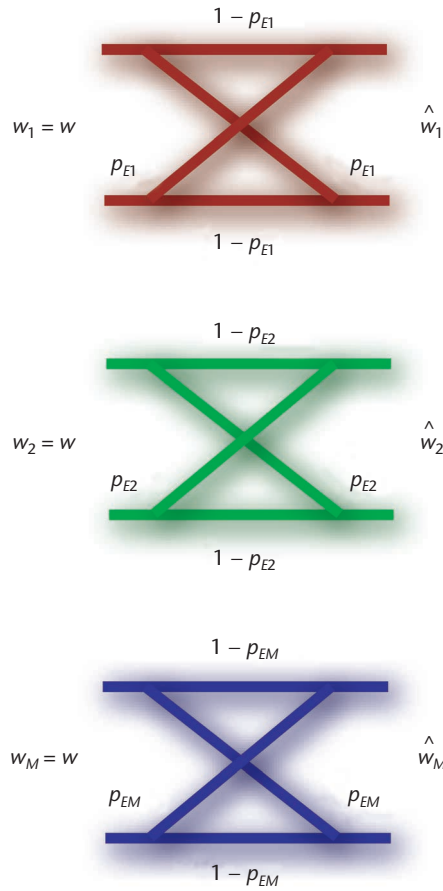
tion. Examples of diversity in wireless communications include antenna and frequency redundancy. For watermarking, we introduce diversity by repeatedly embedding the watermark through the host signal. We call this *coefficient diversity* because we modulate different coefficients within the host signal with the same information.

The sacrifice in employing diversity is the bit-rate expense because the same information is sent through M orthogonal resources. However, for many watermarking applications, the payload is small. In addition, for video watermarking there exists an abundance of data in which to embed the information. Two questions naturally arise: How do we combine the different repetitions of the watermark to maximize the overall performance? In what type of domain should we introduce coefficient diversity to maximize reliability? I'll attempt to answer these questions in the remainder of this article.

Channel estimation

The estimation of a transmitted signal is only as good as the accuracy of the communications channel's model. We incorporate elementary characterizations of the watermark degradations, called *attack characterization*, to demonstrate the importance of postprocessing for watermark detection. Channel coding is useful when we know the channel's characteristics. However, since in the watermarking problem the nature of

Figure 2. Parallel binary symmetric channel model. We can view the binary watermark communication process using diversity as data transmission through a family of binary symmetric channels.



attacks is somewhat unpredictable, we exploit the fact that the watermark is extracted after an attack to maximize extraction reliability.

More needles in the haystack

We can incorporate diversity and channel estimation into our analysis framework through watermark repetition and attack characterization. To formulate our problem more analytically, let's assume that

- the watermark w is binary and of length N_w bits;
- the watermark information is repeatedly embedded $M \geq 1$ times within the host signal;
- each embedded watermark repetition is extracted separately to produce M estimates of the watermark $\hat{w}_1, \hat{w}_2, \dots, \hat{w}_M$; and
- each extracted watermark estimate $\hat{w}_k$ has a known probability of bit error p_{Ek} .

Many proposed watermarking algorithms^{2,8,9} are encompassed by this class of techniques or can be easily modified to fit this category. The specific details of the data embedding and extraction processes aren't relevant. Although we restrict the watermark to a bit sequence and the reliability measure to the bit-error rate, my colleagues and I believe the spirit of the results discussed here holds for nonbinary watermarks with a different reliability measure such as the signal-to-noise ratio (SNR). In fact, recent work¹⁰ demonstrates how we can use diversity with non-bit-error reliability measures.

Our framework is analogous to transmitting the watermark simultaneously along M independent binary symmetric channels (BSC) as Figure 2 shows. We assume the error probabilities $0 \leq p_{Ek} \leq 0.5$ are known and independent of one another. This type of localized characterization of the distortion in the watermark domain allows better modeling of nonstationary fading distortions. Theoretic work assuming stationary watermark channel models² can preclude the benefits that diversity can provide and might limit understanding into the advantages of using one watermarking domain over another.

Greater than the sum of its parts

To estimate the embedded watermark w , we linearly weigh and add the extracted repetitions so that the overall estimate of the i th watermark bit is given by

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

for each watermark bit $i = 1, 2, \dots, N_w$, where $\text{round}[\cdot]$ is the integer round operator. Kundur and Hatzinakos⁵ have shown that

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

minimizes the bit-error rate of the overall extracted watermark estimate $\hat{w}(i)$.

This linear estimation procedure isn't the only alternative for combining the various extracted repetitions, but it's computationally simple and has been successfully implemented and tested.^{5,6} As I'll show through simulations, employing the scale factor α_k provides significant improvement in performance for certain distortions.

Bit-error analysis for this linear estimation appears elsewhere.^{6,7} This article summarizes a general result and focus on implications and insights. A measure of bit error in the estimated watermark⁶ is given by

$$E' \triangleq \sum_{k=1}^M \alpha_k w_k(i) \oplus \hat{w}(i)$$

By analyzing the statistical expectation of E' , denoted $\varepsilon\{E'\}$, we can gain insight into strategies to enhance performance. We can establish the following error statistic bound:⁶

$$\varepsilon\{E'\} \leq \frac{\bar{p}_E}{1-\bar{p}_E} \left[1 - \frac{D(q_a \| q_b)}{\log\left(\frac{1-\bar{p}_E}{\bar{p}_E}\right)} \right]$$

where

$$\bar{p}_E = \frac{1}{M} \sum_{k=1}^M p_{Ek}$$

is the average bit-error rate, and $D(q_a \| q_b)$ is the relative entropy¹¹

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

where the arguments are $q_a(k) = p_{Ek}/(M\bar{p}_E)$ and $q_b(k) = (1-p_{Ek})/(M(1-\bar{p}_E))$. We can see that q_a and q_b are probability-like distributions because their elements are nonnegative and sum to one. Kundur and Hatzinakos⁶ explain that the error bound is tight for $\bar{p}_E$ and p_{Ek} close to a constant. The equality of the error bound holds if and only if $p_{Ek} = 0$ for all k . A smaller value for the bound on $\varepsilon\{E'\}$ implies that, for the most part, we can guarantee better accuracy of the extracted watermark and, hence, greater robustness.

Implications and design insights

From our analysis result, we find that the following possible tactics may be incorporated into a watermarking scheme to lower the value of the error statistic bound on $\varepsilon\{E'\}$ and, hence, improve the watermarking system's robustness in some way.

Scale down the haystack. Reducing the average bit-error probability $\bar{p}_E$ decreases the term $(\bar{p}_E/(1-\bar{p}_E))$ and increases the denominator term $\log((1-\bar{p}_E)/\bar{p}_E)$. Both lower the overall bound. Many

proposed watermarking methods attempt to gain performance by diminishing this average error rate. These methods commonly employ signal-processing strategies to imperceptibly embed a higher energy and, on average, a more robust watermark. The deficiency of many watermarking methods is that they solely rely on embedding a stronger watermark using sophisticated human perceptual mathematical models for improved performance. The next two theoretical observations shed light on a different strategy to increase robustness.

Shape up the haystack. Specifically, embed the watermark such that the distributions q_a and q_b are dissimilar for a large class of distortions. For a fixed value of $\bar{p}_E$, we can reduce the performance bound by increasing the value of $D(q_a \| q_b)$. The relative entropy is a measure of the distance between its two argument distributions.¹¹ Roughly, we can see that $D(q_a \| q_b)$ is large when $q_a(k) = p_{Ek}/(M\bar{p}_E)$ and its corresponding $q_b(k) = (1-p_{Ek})/(M(1-\bar{p}_E))$ are dissimilar.

Assuming a fixed average probability of bit error, this requires that p_{Ek} vary in amplitude for different values of k . This implies that we should embed the watermark in a domain for which the degree of distortion varies in each localized region containing a repetition of the watermark. As a result, the amplitude of p_{Ek} will differ for distinct values of k . We can achieve this by inserting the watermark in a domain that distributes the distortion more to certain coefficients, leaving the others less affected.

Uncover some needles. Localizing the distortions on the watermarked signal enhances robustness. Kundur⁷ showed that the existence of $p_{Ek} = 0$ for at least one $k \in \{1, 2, \dots, M\}$ implies that $E' = 0$. Thus, if a set of localized coefficients containing one complete repetition of the watermark unmodified by the distortion exists, then perfect watermark recovery is possible as long as we know all the values of p_{Ek} . This translates to embedding the watermark in a domain that can completely localize the distortion to a finite and relatively small percent of the coefficients.

The first two implications relate the accuracy of the extracted watermark to the watermark domain in which the hidden data is embedded. By using diversity and attack characterization, we can improve the watermark's effectiveness to a specific class of distortions by inserting the mark in signal coefficients that localize these distortions. For example, to design a watermark that is

robust against cropping, it would be wise to embed the mark in the spatial domain, which completely localizes the manipulation. Although a portion of the watermark is clipped out, the repetitions in the remaining signal are still accessible. Similarly, for robustness against filtering, we should embed the watermark in the discrete Fourier domain, which localizes the associated degradations. Mild linear filtering will affect some Fourier coefficients more than others. To make the watermark robust to both, a compromise is to use the discrete wavelet domain.

Kundur and Hatzinakos¹² present more specific work on incorporating particular wavelets to analyze their robustness to perceptual coding in various multiresolution domains. They demonstrated that using complementary domains for watermarking and perceptual coding improves the robustness of the embedded watermark. This work is in direct conflict with well-established principles that suggest the same domain is superior.¹³ Recent theoretical work¹⁴ attempts to determine appropriate domains for watermarking in the face of perceptual coding.

Practical application and results

To verify the insights derived in this work, we implemented the proposed principles in a practical watermarking method called robust reference watermarking.^{5,6}

Robust reference watermarking

Our implementation embeds the watermark in the wavelet domain that localizes these degradations because of robustness against frequency and spatial domain distortions. (I won't go into the specific details of the method here but refer readers to the relevant literature.^{5,6})

The scheme essentially embeds and extracts two types of binary watermarks in a given watermarking domain (for our implementation it's the wavelet domain). The first type of watermark, the robust watermark, contains the payload. The second type, the reference watermark, estimates the robust watermark's reliability. Both of these marks are repeatedly embedded in localized regions of coefficients at each resolution. We extracted the different repetitions of the robust and reference watermark separately. The reference watermark repetitions are each used to estimate the probability of bit error of the robust watermark's closest repetition. We used the linear weighting I discussed in this article to find an overall optimal linear estimate of the embedded robust watermark.

Based on our analysis, we believe that the strengths of robust reference watermarking arise from the following factors:

- The watermark is embedded in the wavelet domain, which localizes a diverse class of common signal distortions such as cropping and frequency domain filtering.
- The reference watermark characterizes the locally varying distortions on the watermarked signal before extraction. In this way, the watermark's components are more accurate than the others.
- The reference watermark provides an objective measure of each embedded watermark repetition's reliability. The robust watermark isn't used for estimation of its own reliability because this results in greater robustness but increases the false-positive error-detection rate.⁷
- The proposed scheme is robust to a wider variety of distortions because we employed diversity strategies by using watermark repetition. Any single repetition of the watermark needs to be intact for reliable watermark recovery.

Testing

To demonstrate the advantages of diversity and channel estimation, we compared the performance of the robust reference watermarking technique to full characterization using the reference watermark and without any reference watermark, assuming the reliability of each repetition is the same. We used the correlation coefficient between the embedded and extracted watermarks as the measure of robustness.

In the implementation of this method, we specifically used Daubechies 10-point wavelet for all simulations. We used the parameter values $L = 4$ and $Q = 3$ for the robust reference watermarking algorithm.⁶ The robust and reference watermarks are randomly generated, binary, and 128 bits in length. We performed the simulations on a 256×256 host image (see Figure 3a). Figure 3b shows the watermarked image. The remainder of the results are presented as plots of correlation coefficient (robustness) versus degree of attack. The solid lines represent the performance without attack characterization. The dashed lines are with attack characterization to demonstrate the improvement when employing effective diversity.

We added white Gaussian noise to the watermarked image to determine the methods' robust-

ness to stationary additive interference. Figure 4a presents the results. Visible image degradation was apparent around an SNR of 30 decibels. The watermark, however, has a high correlation for even higher noise levels. As the plots demonstrate, we can achieve some improvement with diversity.

Figure 4b shows the performance improvement for general lowpass filtering. A radially symmetric blurring function of the form $ca^{m^2+n^2}$, where m and n are the spatial coordinates, c is a normalization constant, and a is a parameter that dictates the degree of filtering is applied to the watermark signal before watermark extraction. We can achieve similar enhancement for nonlinear median filtering (not shown). Figure 4c shows the JPEG compression results for different compression ratios.

One of the most effective attacks against watermarking schemes proposed by Barnett and Pearson¹⁵ is called the laplacian removal attack operator. The technique attempts to estimate the watermark from the watermarked signal's high-frequency components and then subtract it out of the watermarked signal. The relative degree of watermark removal is provided by a parameter used in our sim-



Figure 3. Test example.
(a) Original host image, and (b) resulting watermarked image used in simulations.

ulations. Figure 4d shows the results for this sophisticated attack. Note that the watermark's presence is apparent even when the Laplacian removal attack significantly degrades performance.

Final remarks

This article demonstrates that watermark repetition when combined with attack characterization is a powerful approach to improve and broaden robust watermarking performance. By assuming a

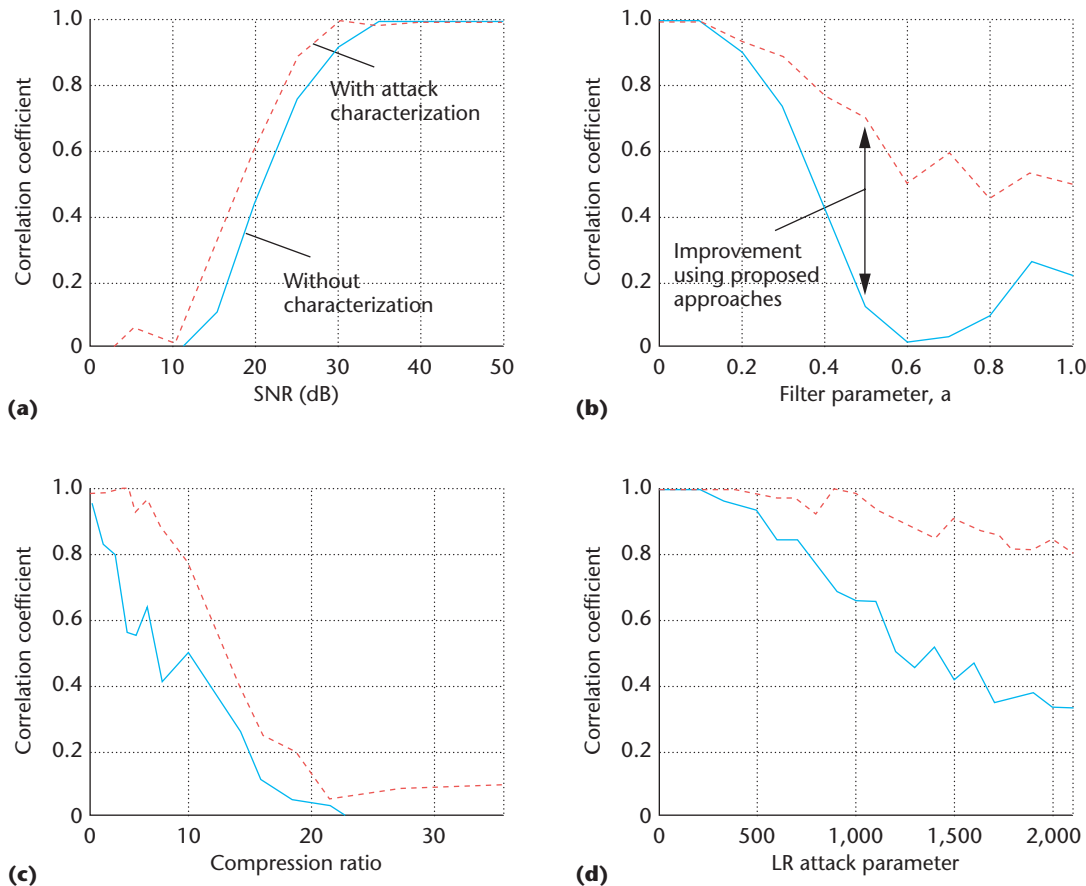


Figure 4. Summary of tests showing improvement that we can achieve by employing diversity effectively in a watermarking system.

localized nonuniform degradation model for the watermark we gain insight into appropriate domains in which to robustly hide data. It's evident from the results that diversity and channel estimation improves the absolute performance of a robust data embedding method for a common class of attacks and expands the class of distortions for which the watermark is resilient.

Future work involves investigating the use of novel tool sets such as space-time coding for robust watermarking to more effectively distribute our needles in multiple haystacks. **MM**

References

1. R.G. van Schyndel, A.Z. Tirkel, and C.F. Osborne, "A Digital Watermark," *Proc. IEEE Int'l Conf. Image Processing*, vol. 2, IEEE CS Press, Los Alamitos, Calif., 1994, pp. 86-90.
2. I.J. Cox et al., *Secure Spread Spectrum Watermarking for Multimedia*, tech. rep. 95-10, NEC Research Inst., Princeton, N.J., 1995.
3. R.B. Wolfgang and E.J. Delp, "A Watermark for Digital Images," *Proc. IEEE Int'l Conf. Image Processing*, vol. 3, IEEE CS Press, Los Alamitos, Calif., 1996, pp. 219-222.
4. P.G. Flikkema, "Spread-Spectrum Techniques for Wireless Communications," *IEEE Signal Processing*, vol. 14, no. 3, May 1997, pp. 26-36.
5. D. Kundur and D. Hatzinakos, "Improved Robust Watermarking through Attack Characterization," *Optics Express Focus*, vol. 3, no. 12, 7 Dec. 1998, pp. 485-490.
6. D. Kundur and D. Hatzinakos, "Diversity and Attack Characterization for Improved Robust Watermarking," to be published in *IEEE Trans. Signal Processing*, vol. 49, no. 10, Oct. 2001.
7. D. Kundur, *Multiresolution Digital Watermarking: Algorithms and Implications for Multimedia Signals*, doctoral thesis, Univ. of Toronto, 1999, <http://www.comm.toronto.edu/~deepa/pub.html>.
8. C.I. Podilchuk and W. Zeng, "Image-Adaptive Watermarking Using Visual Models," *IEEE J. Selected Areas in Communications*, vol. 16, no. 4, May 1998, pp. 525-539.
9. J.J.K. Ruanaidh and T. Pun, "Rotation, Scale and Translation Invariant Digital Image Watermarking," *Proc. IEEE Int'l Conf. Image Processing*, vol. 1, IEEE CS Press, Los Alamitos, Calif., 1997, pp. 536-539.
10. S. Voloshynovskiy et al., "Optimal Adaptive Diversity Watermarking with Channel State Estimation," *Proc. SPIE Security and Watermarking of Multimedia Contents III*, vol. 4134, SPIE Press, Bellingham, Wa., 2001, pp. 23-27.
11. T. Cover and J. Thomas, *Elements of Information Theory*, John Wiley & Sons, New York, 1991.
12. D. Kundur and D. Hatzinakos, "Mismatching Perceptual Models for Effective Watermarking in the Presence of Compression," *Proc. SPIE Multimedia Systems and Applications II*, vol. 3845, SPIE Press, Bellingham, Wa., 1999, pp. 29-42.
13. R.B. Wolfgang, C.I. Podilchuk, and E.J. Delp, "The Effect of Matching Watermark and Compression Transforms Incompressed Color Images," *Proc. IEEE Int'l Conf. Image Processing*, vol. 1, IEEE CS Press, Los Alamitos, Calif., 1998, pp. 440-444.
14. C. Fei, D. Kundur, and R. Kwong, "The Choice of Watermark Domain in the Presence of Compression," *Proc. IEEE Int'l Conf. Information Technology: Coding and Computing*, IEEE CS Press, Los Alamitos, Calif., 2001, pp. 79-84.
15. R. Barnett and D.E. Pearson, "Attack Operators for Digitally Watermarked Images," *IEEE Proceedings, Visual Image and Signal Processing*, vol. 145, no. 4, Aug. 1998, pp. 271-279.



Deepa Kundur is an assistant professor in the Edward S. Rogers Sr. Department of Electrical and Computer Engineering at the University of Toronto. She is the Bell Canada Junior Chair-holder in

multimedia and is an associate of the Nortel Institute for Telecommunications. She received her BS, MS, and PhD from the Department of Electrical and Computer Engineering at the University of Toronto. She has written several papers in the content-based security area and has helped organize special sessions on multimedia security at various conferences. Her research interests span the areas of multimedia security, data hiding and covert communications, content-based multimedia processing, and nonlinear and adaptive communication algorithms. She is a member of the IEEE Communications and IEEE Signal Processing societies and the Professional Engineers of Ontario (PEO).

Readers may contact Kundur at the Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, 10 King's College Rd., Toronto, Ontario, Canada, M5S 3G4, email deepa@comm.toronto.edu.

For further information on this or any other computing topic, please visit our Digital Library at <http://computer.org/publications/dlib>.