Communication and Information Theory in Watermarking: A Survey

Adrian Sequeira and Deepa Kundur

Edward S. Rogers Sr. Department of Electrical and Computer Engineering
University of Toronto, Toronto, Ontario, Canada M5S 3G4

ABSTRACT

This paper presents a review of some influential work in the area of digital watermarking using communications and information-theoretic analysis. After a brief introduction, some popular approaches are classified into different groups and an overview of various algorithms and analysis is provided. Insights and potential future trends in the area of watermarking theory are discussed.

Keywords: watermarking theory, data hiding models, information theory, communication theory, survey.

1. INTRODUCTION

Digital watermarking is the process of discreetly and robustly embedding information, called a *watermark*, in media signals to provide some form of added-value; applications include broadcast monitoring, signal tagging, and copy control. The embedding process involves imperceptibly modifying a *cover* media signal using a secret key and the watermark to produce a composite *watermarked* signal. The modifications are made such that reliable (i.e., robust) extraction of the of the embedded watermark using the secret key is possible even under a "reasonable" level of distortion applied to the watermarked signal. These distortions, whether intentional or incidental, are known as *attacks*.

As the research area of digital watermarking matures, one can see some general trends in its development. There was initial work in the use of basic digital signal processing (DSP) strategies for data hiding. Robustness-enhancing strategies were employed using intuition on human perception and basic communications. However, as the area has grown, some theory is emerging. This framework aims to unify much of the past work and establish technical insights for future algorithms. The new mathematical language for describing watermarking borrows tools from statistical communications and information theory.

The purpose of this paper is to review some recent work in the area of watermarking theory. Our main objectives are:

1. to provide perspective regarding the role of communication and information theories in digital watermarking, and
2. to highlight some of the trends in the area.

The art of digital watermarking involves the judicious selection of technological trade-offs to develop an algorithm suitable for a particular application. There are many different factors involved in determining an appropriate compromise; these include cryptographic security, psychology of perception, robustness of extraction, statistical false extraction rates, and complexity. The communication and information theoretic approaches to analysis primarily address the interplay between watermark robustness, capacity and signal strength. These issues will, therefore, also be the central focus of this paper.

Digital watermarking is analogous to the problem of reliable communications as shown in Figure 1. The process of conveying watermark information through a media signal in the face of attacks is equivalent to communications in a hostile environment. The process of watermark embedding is analogous to channel coding, watermark extraction serves the role of a communications receiver, and the effective communication channel is characterized by the nature of attacks applied to the watermarked signal. Depending on the design of the system, the watermark may be error

Email: {adrian, deepa}@comm.toronto.edu

Multimedia Systems and Applications IV, Andrew G. Tescher, Bhaskaran Vasudev, V. Michael Bove, Jr., Editors, Proceedings of SPIE Vol. 4518 (2001) © 2001 SPIE · 0277-786X/01/$15.00

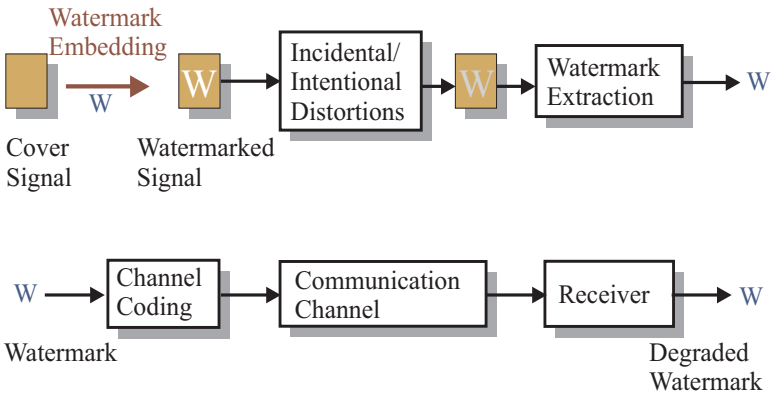

Figure 1. Watermarking as a Communications Problem.

correction coded before embedding into the cover data. The cover medium may be treated as the channel noise or as some (partial) channel state information. The attacks in the equivalent channel are often of unknown statistics in which case the detector used in watermarking can be designed to use estimated or assumed channel statistics.

Many of the tools used to improve communications may also be employed to promote more robust watermarking. It is convenient to apply traditional theories developed for communications to form a sort of theory for watermarking. Although this theory primarily focuses on the robustness component of the watermarking problem, and is therefore, incomplete, it is, nevertheless, useful for many aspects of algorithm design and assessment.

In the next section, we propose our classification scheme of the literature in watermarking theory. Sections 3 and 4 provide a survey of communication theory and information theory in watermark, respectively. The paper concludes with a discussion of themes and trends in the area.

2. CLASSIFICATION

Communication theory approaches are more practical than their information-theoretic counterparts and lend themselves more easily to algorithm design. Existing work in the area of watermarking theory may be classified into two groups: watermarking based on communication theory, and watermarking based on information theory. There is some overlap between the two areas; we distinguish the work by defining communication theory-based techniques as those that use specific communication theory tools such as statistical detector analysis to analyze and design specific watermarking algorithms. In contrast, information-theoretic watermarking uses more general analysis to derive ultimate performance bounds and optimal coding strategies for a particular watermarking approach, or a specific attack class. In communication theory-based watermarking, robustness is often measured by the correlation coefficient between the embedded and extracted watermark or bit error rate. For information-theoretic watermarking, it is assessed through data hiding capacity, the theoretically maximum number of bits that can be reliably embedded and extracted from the media signal.

Both classes of watermarking theory research can be subdivided into smaller categories as discussed in the subsequent sections. Figure 2 provides an overview of our proposed classification.

2.1. Communication Theory-Based Approaches

We divide communication theory-based watermarking into the following three groups:

1. Algorithm Development: these papers propose a new algorithm or modify an existing one and test it. They look at the communication system as a whole and describe, develop and integrate all the components for transmission and reception of the watermark signal.

2. Attack Analysis: these approaches concentrate on qualifying and quantifying attacks and their effects, and involve developing countermeasures against them. These papers focus on the channel in the communication system and attempt to develop new techniques of embedding or detecting the watermark.

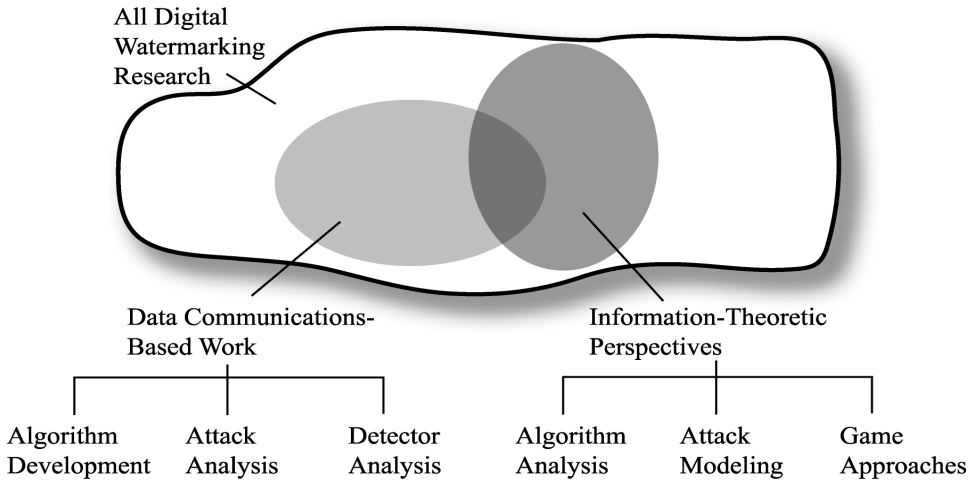

Figure 2. Classification of Research in the Area of Watermarking Theory.

3. Detector Analysis: these strategies look at improving the receivers in a watermarking channel. The detectors are built based on the noise assumptions of the channel. The most popular are correlator detectors which are ideal for the white noise assumption. Other designs have also been developed for different noise models.

2.2. Information Theoretic Approaches

Information theoretic approaches are concerned with bounds and ultimate watermarking system trade-offs. They are somewhat more general in implication than their data communication counterparts. The methods under this class can be classified into the following three groups:

1. Algorithm Analysis: the approaches in this class consider a given digital watermarking system, model the various signals, processing and attacks, and derive bounds regarding the ultimate robustness of the specific technique.

2. Attack Modeling: these strategies essentially consider a given type of watermark attack and based on this measure attempt to derive optimal embedding and detection strategies to maximize the overall capacity.

3. Game Approaches: in these techniques the entire watermarking process is considered to be a game and optimal joint data hiding and attack strategies are derived. The watermarker and attacker are both assumed to be at "peak" performance. Expressions for the possible data hiding rate are derived and analyzed to gain perspective as to the potential of the technology.

3. COMMUNICATION THEORY IN WATERMARKING

The most common watermark requirements are invisibility and robustness. These requirements are somewhat contradictory to each other as invisibility implies the watermark signal has low power in the media to avoid perception, while robustness requires the watermark to have large power to help statistical detection. Under these conditions, spread-spectrum (SS) communication theory initially emerged as appropriate for watermarking because it spreads the watermark signal with a low amplitude but in a wide enough bandwidth to hold enough signal energy for detection.

However, there is another caveat. A SS signal is usually designed to be pseudo random in nature with an autocorrelation resembling that of white signals. The best detector is a matched filter correlator. This works well in a regular communication channel where the signal energy is a fair amount higher than noise energy. However in a watermarking channel, because the watermark is at a significantly lower power than the signal, there is quite a bit of interference from the cover data which is the channel for the watermark signal.

In addition, there is the threat of incidental or intentional attacks that can remove or render watermarks undetectable. Examples of incidental attacks are image/audio/video compression or packet loss. Intentional distortions

include filtering, rotation, scaling, or manipulations which exploit a particular class of embedding algorithms. The case of compression is particularly interesting as it is one of the most popular attacks studied in the literature; in fact, if perfect lossy compression existed, the data hiding problem would be impossible[1]! It works by removing the high frequency components of the image, which are the least perceptually significant, resulting in little or no loss in image quality. Thus, any watermark placed in these locations will be definitely lost.

3.1. Algorithm Development

Cox *et al.*[2] were the one of the first to develop a technique combining these ideas. They used a watermark distributed as zero-mean unit variance Gaussian of sufficiently low power embedded in the 1000 lowest (non- DC) DCT coefficients of the images. Detection of the watermark was by using a similarity measure, similar to a correlator detector.

Many authors since have used the SS concept, although in different ways: Piva *et al.*[3] also embedded the watermark in the DCT domain as did Cox; Kohda *et al.*[4] used the DCT of the YIQ encoding of an image; Wong *et al.*, log-2 spatio domain[5]; Checcacci *et al.*[6] developed a replicative method for embedding in a compressed MPEG-4 video stream.

In general, all these techniques can be represented in Figure 3. To embed the watermark, the cover media and the watermark may, if necessary, be transformed into a suitable domain and combined in an embedder using a key K. The key is used to provide secrecy either by randomizing the positions of the pulses in the medium or by generating appropriate pulses. The most common technique is the addition of the watermark to the cover signal, after the watermark has been suitably spread. The resulting watermarked medium is then inverse transformed to give the watermarked transmittable medium. To extract the watermark, the received and possibly attacked medium is transformed if necessary. This is then sent to a detector and usually only if the watermark is present is it extracted for decoding. The most common detector design is the matched filter, implemented as a correlator, with thresholding of the results.

With the development of many algorithms, the need for a theoretical analysis arose. A theoretical analysis would help improve watermarking techniques by setting the specific conditions for future algorithms. Cox *et al.*,[7,8] developed a relatively new technique based on the idea of treating the watermarking channel as a system of communications with side information. The idea was that instead of treating the cover data as noise added to the medium, it could be treated as side information and used to improve the fidelity and detectability characteristics of the watermark by means of an appropriate perceptual mask and knowledge of the detector. This side information is extracted by a function and, for this work, is assumed zero-mean independent and Gaussian after extraction. The watermark is then *mixed* with this side information before embedding back into the original media to produce the watermarked medium ready for transmission.

In more recent work, Voloshynovskiy *et al.*,[9–11] have come up with a multi-level watermark embedding scheme using side information only at the decoder. This side information is in the form of the image statistics made available to the detector in its design. Two cover image statistics are assumed and tested based on past experience[12,13]; in particular, the locally non-stationary Gaussian and the global stationary generalized Gaussian distributions are employed. The detectors are matched filters based on these assumptions. In addition to an additive SS-based watermark, the algorithm includes a key based reference watermark for synchronization and channel estimation, After thresholding the matched filter output, a maximum likelihood (ML) or maximum a posteriori decoder (MAP) is used to estimate the watermark message.

3.2. Attack Analysis Approaches

Continuing with the analogy of watermarking as a communications system, some researchers have chosen to work on modeling and resisting attacks on the watermark. They work on the philosophy that the more specific the information known about the family of possible attacks, the better we can design systems to resist it.

Kundur *et al.*[14] assert that certain attacks such as cropping, filtering and perceptual coding can be modeled as fading in a noise attenuating non-stationary channel. Thus, they employ principles of diversity and channel estimation to improve performance of watermarking schemes. Analysis is provided to show that for common attacks such as spatial cropping and compression, the wavelet-domain, which tends to isolate these distortions, is one of the best domains in which to embed the information. The approach is implemented in a technique known as Robust Reference Watermarking (RRW) which employs watermark repetition and a reference watermark to estimate the

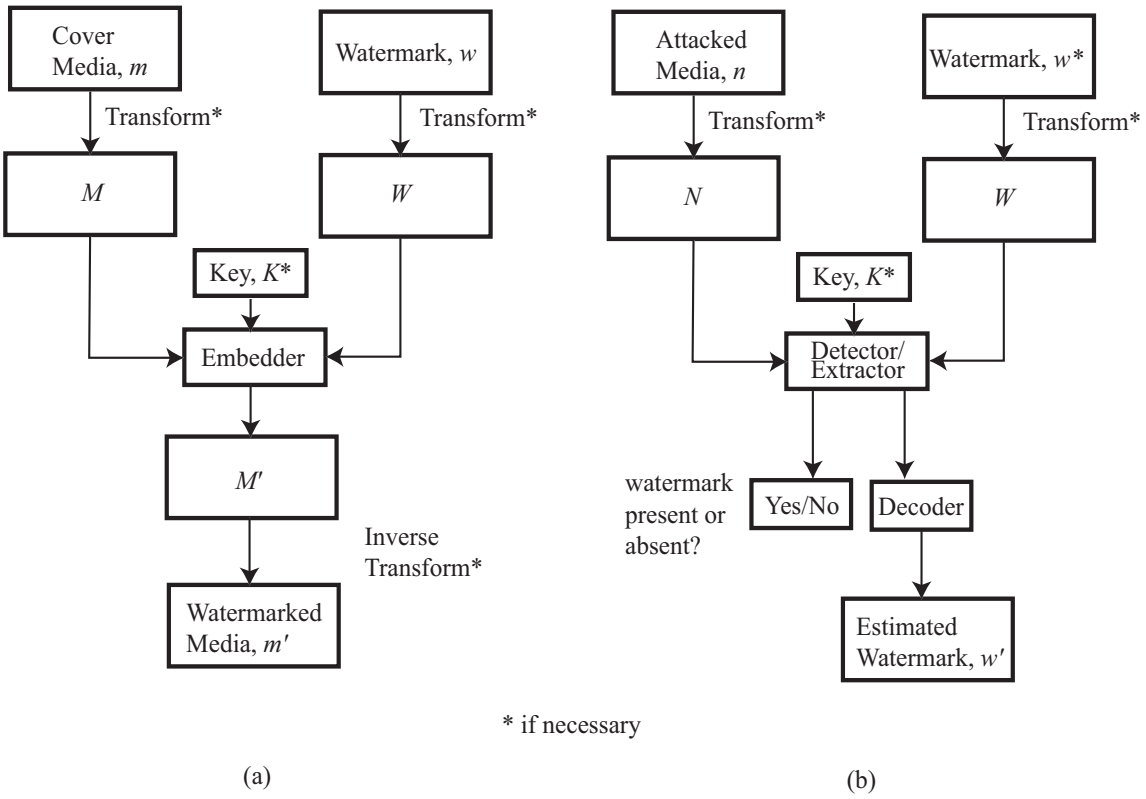

Figure 3. General Architecture for Algorithm Development Approaches: (a) General Embedding Process, (b) General Decoding/Detecting Process.

attack characteristics. Simulation results verify their observations demonstrating that the class of attacks for which a watermarking scheme is robust can be greatly broadened.

Su and Girod[15–17] analyze the Wiener estimation attack and develop the concept of energy efficient watermarking. Wiener estimation is a power spectral density-based linear filtering of the received watermarked data to either estimate the cover data or the watermark itself. Estimation of the cover data is useful for improving detector performance for *blind* watermarking*.[18] The Wiener estimation attack involves estimating the watermark before removing it. To combat this, Su and Girod developed the idea of energy efficient watermarking, i.e., the lower the amount of watermark energy that can be estimated, the smaller the amount that can be removed. This requires a large error between the watermark estimate used in the attack and the actual embedded watermark itself, and is achieved if the watermark power spectral density has the same shape as the cover data. Such a watermark is known as power spectrum compliant (PSC).

Su and Girod[17,16] also show that the Wiener attack is optimal in the sense that it produces the least attack energy distortion for a given correlator detector output. However, if the optimal receiver is used, which is the correlator for Gaussian noise, then PSC watermarks prove to perform better at high distortions while white watermarks perform better for low distortions under the Wiener attack. The PSC is also the best defense against white noise attacks, but at high distortion levels.

Linnartz *et al.*[19,20] analyze another type of attack against binary watermarks: the sensitivity attack. The authors assume that the detector is standardized and readily available to an attacker. Since most detectors decide on the presence or absence of the watermark by comparing the correlation value with a threshold, the authors decide to find the threshold using test images differing from each other by changes in luminance of a few pixels.

*Blind watermarking refers to extraction of the mark without knowledge of the cover media used in the initial embedding. Non-blind watermarking, thus, involves extraction using the cover media.

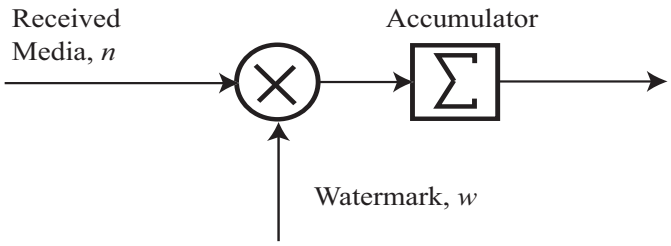

Figure 4. Blind Correlation Receiver Structure for Watermarking.

Using the detector, the image space can be divided into those signals that give correlator outputs less than the threshold and those that are greater than the threshold. So the attacker can find the tangent to the curve that divides these two regions, which involves testing all pixel positions. Now that the tangent is known, the attacker can subtract just enough out of the luminance of the original image to give a negative detection result and very small perceptual difference. This process can be iterated if the attacker is not satisfied with the perceptual damage.

As a solution, the authors proposed using a decision interval instead of a decision threshold value. This interval would work on making the threshold less predictable as any correlation below/above the lower/upper limits of the interval would mean absence/presence of the watermark. Values in the interval would be randomly 0 or 1 based on a pdf of optimal shape (squared sine function, in particular) to reduce the information leakage. It is found that the larger the decision interval, the lower the information leakage and the lower reliability of the detector, i.e., higher probabilities of false alarm and missed detection. The authors conclude that the crux of the problem with correlator watermark detectors is their linear nature.

3.3. Detector Analysis Work

In the development of digital communication systems, the optimal detector to maximize the signal-to-noise ratio (SNR) in the face of stationary additive white noise is the matched filter detector. Over time, refinements to this basic design have ensured low error rates in communication. In a similar vein, knowledge of the watermark channel characteristic allows for better detectors and decoders to be built. Unfortunately, watermark channel characteristics are not known precisely. Hence novel detector and decoder design is still under development in current literature.

Linnartz *et al.*[21] attempt to develop criteria for reliability of watermark detection. The authors make no prior assumptions for statistics of the cover image or the watermark, but estimate them from the image and watermark themselves. Both white and low pass watermarks were tested with a correlator detector (shown in Figure 4) and it was found that low pass watermarks have a lower reliability because there is more interference from the cover image itself. Depovere *et al.*,[22] note that adding a whitening filter in the detector before both the cover image and watermark are correlated can improve reliability considerably.

In order to obtain more general results, in,[23,24] the authors analyze the effect of spatial correlation and DC offsets in watermark detection. In this case, the authors analyze a quasi white watermark (i.e., a white watermark with a significant DC component) and a low pass watermark, using a correlator detector. They note that the presence of offsets and of spatial correlation can affect the reliability of the detector by affecting the false positive rates.

In Ref. 18, Hernandez *et al.* develop a 2D multipulse amplitude modulated scheme and analyze the detector. The authors propose an embedding scenario in which the image is divided into orthogonal regions each with a 2D pulse of different strengths modulated by an encoded watermark. In Ref. 25, it is found that the best code for performance, measured in bit error rate (BER), is the one with the least redundancy, i.e., the one with the longest message length for codewords of the same length. This scheme models the channel as a Gaussian vector channel based on the idea that the correlation coefficients for each bit of a codeword are Gaussian distributed; the cross correlation values are considered negligible. This model is also applied to attacks modeled as linear filtering for a maximum likelihood (ML) detector.

In Ref. 12,13, Hernandez *et al.* analyze an additive SS watermarking system from the point of view of developing an efficient detector. They analyze for watermarks in two domains: spatial and discrete cosine transform (DCT); the same embedding technique is used as in Ref. 18 and a correlator detector is employed. The statistics of the correlation values are assumed Gaussian distributed. The authors argue that the optimal detector in the spatial

domain is a log-likelihood function and that for the DCT is the Neyman-Pearson[26] detector, while in both domains the optimal decoder is the ML decoder. The ML decoder is ideal for Additive White Gaussian Noise (AWGN) channels since it matches the received signal to the closest point in the Euclidean distance sense. The Neyman-Pearson on the other hand works to minimize the probability of missed detection while keeping the probability of false alarm constant. This works well especially for non-Gaussian distributions. For the spatial domain, the statistics are estimated and assumed to be quasi-stationary and ergodic; quasi-stationarity is necessary since there is some correlation between adjacent pixels and ergodicity is needed so that the first and second moments can be easily estimated for detection and decoding. For the DCT domain, the non-DC coefficients are assumed to fall into a generalized Gaussian distribution.[27] The spatial domain detector gives results matching well with theory under different attacks, while the DCT domain detector was found to give the worst results for Gaussian noise statistics.

Other publications[28,29] have dealt with analyzing multiplicative watermarks by assuming the Weibull detection statistics or using a Bayesian approach to detection (i.e., Gaussian statistics assumed). We refer the interested reader to the specific references for more detail.

4. INFORMATION THEORY IN WATERMARKING

The information theoretic work involves, for the most part, more general analysis of the watermarking problem. We discuss in this section the different classes and summarize some influential work in the area. The area is rapidly developing; at the time of this writing a new special issue committed to information-theoretic watermarking approaches is to be published.[30] As we focus on work already published and, therefore, more readily available, we refer the reader to Ref. 30 for more information.

4.1. Algorithm Analysis Approaches

The techniques in this class are concerned with finding performance bounds of a particular type of watermarking scheme. Specifically, in this section *capacity* refers to the greatest number of bits that can be embedded using a specific watermark embedding scheme and for a particular class of attacks.

Some earlier work focused on employing information theory to determine ultimate performance bounds for well-established approaches in the literature. In Ref. 31, the Barni *et al.* propose a numerically-based capacity estimation procedure for a form of blind spread-spectrum watermarking in the DCT-domain for which the primary source of distortion is the cover signal interference. The effect of any attacks on the watermark capacity are not taken into account in the work. Specifically, a discrete memoryless channel (DMC) model is employed for the cover signal interference. The DCT cover signal coefficients are modeled using a generalized Gaussian pdf which does not facilitate a closed-form expression for the overall data hiding capacity. Numerical estimation procedures are used to derive capacity bounds on specific test images. The authors extend the work in Ref. 32 by including capacity results for spread-spectrum watermarking in the discrete Fourier transform (DFT) magnitude domain. The DFT coefficient magnitudes are modeled using a Weibull distribution. By ignoring the issue of different masking properties in the DCT and DFT domains and keeping the watermark strength uniform for all coefficients in both domains, the DFT appears to have a higher capacity than the DCT. However, as the authors point out, the diverse masking properties will provide a more accurate understanding of the capacity of each domain and the superiority of one over the other.

Chen and Wornell[33,34] introduce a class of watermarking methods called quantization index modulation (QIM) for data hiding. These class of techniques embed the watermark in a cover signal through quantization; a different quantization vector is used to embed a different watermark value. They analyze this quantization-based watermarking approach through capacity analysis. A framework is developed for analyzing achievable performance trade-offs for robustness, distortion and embedding rate. Their work involves the application of coding theory measures to different classes of watermarking schemes to derive insights regarding ultimate trade-offs among capacity, robustness and imperceptibility. Specifically, through their framework, they equate capacity for watermarking with the number of possible distinct quantizers that can be used for reliable embedding, the transparency to the energy of the watermark (or more generally the shape of the *quantization cells*), and robustness to the minimum distance measures from coding theory. Using a bounded perturbation model for the possible watermark attacks, they compute various information-theoretic trade-offs for QIM, SS and low bit modulation (LBM) (in which, for example, the least significant bit is replaced with the watermark values).

They show that QIM structures are optimal when the watermark is energy-constrained, and the watermark channel is memoryless. Capacity results are derived for various implementations of QIM. In addition, the authors

argue that the distortion an attacker needs to impose in order to remove the watermark is greater for QIM than it is for SS and LBM. Hence, an attacker would be forced to apply more perceptible distortion to remove the mark if QIM was used.

In general, the work in this class sheds initial light on the potential of watermarking technology. However, because the approaches do not involve the psychological element to data hiding such as notions of masking; that is, the issue of imperceptibility is not defined in more sophisticated terms than an energy constraint, the ideas do not provide insight into the dependence of maximum data hiding rate on the cover signal. In the next section, we look into slightly more general approaches that focus on modeling the attack and attempt to derive optimal channel coding strategies for effective watermark communication.

4.2. Attack Modeling

The approaches of this class find ultimate performance bounds assuming a specific attack on the watermarked signal. Capacity refers to the maximum number of watermark bits that can be reliably embedded in the face of a specific attack. Although, the analysis is restricted to a class of attacks, the embedding and extraction routines are not significantly limited. Most of the initial work in this area assumed that the interference on the watermark can be modeled as additive Gaussian noise. Therefore, traditional Gaussian watermarks were considered to be optimal in terms of capacity maximization.

However, when the idea of watermarking was initially proposed, the exact form of attacks for a broad class of applications was unclear. Intentional attacks are more difficult to model accurately for a specific application as they depend on the psychology of the opponent in the problem. Incidental distortions, however, depend on predictable signal manipulations for a given application. One of the most common incidental distortions is compression.

Kundur[35] considers the specific problem of determining and achieving capacity in the face of quantization-based compression; compression and watermarking are assumed to occur in the same domain. Quantization is modeled as a uniformly distributed random variable, and capacity-achieving channel codes are derived. Unlike, traditional Gaussian watermarks that are used in the literature, it is determined that binary watermarks provide the best performance in terms of capacity. In addition, energy allocation principles based on the water-filling problem are employed in order to determine how much of the watermark energy should be embedded in the different cover signal coefficients.

The work is extended to different compression and watermarking domains by Fei *et al.*[36,37] In this work, capacities of SS and quantization-based watermarking[38] for different domains in the face of JPEG compression are derived. It is found that for different compression ratios, different embedding strategies are optimal. Thus, a hybrid watermarking scheme using different embedding strategies for different coefficients in the watermark embedding domain is developed for robustness against JPEG compression.

4.3. Game-Based Analysis

In game-based analysis, no severe restrictions are initially set on the type of data embedding, extraction or attack. Instead it is assumed that both the embedder and the attacker work at "peak" performance depending on the knowledge available to them. Using this formulation, the *watermarking game* becomes an optimization problem in which the watermark embedder wants to maximize data hiding capacity and the attacker intends to minimize this quantity. Hence, a min max problem with respect to various system parameters results.

O'Sullivan *et al.*[39] first formulated the watermarking scenario as a game played between an information hider and an attacker. The cost function in the game is the mutual information between the input and output of the attack channel; the attacker tries to minimize this cost while the hider tries to maximize it. The upper bound on this mutual information is the data hiding capacity. The main result of the work is the insight that the best attack approach is equivalent to the most efficient data compression possible subject to a distortion constraint, and the optimal information hiding strategy corresponds to optimal channel coding in which the attacker determines the channel characteristics. The work is extended in Refs. 40 and 41 in which watermarking is formulated as a communication problem with side information. Their work determines data hiding capacity which quantifies the fundamental compromise among achievable watermark rates, allowable distortions for the embedder and allowable distortion for the attacker. Although no general closed-form expression comes out of the work, results are provided assuming Gaussian host signal.

In related work by the authors, when watermarking is performed specifically in the wavelet domain, Moulin *et al.* model the cover signal coefficients' sparsity using statistical spike models. The main results state that for optimal embedding, the power must be equalized along strong channel components of the cover data[42]; likewise, the optimal attack must do this as well. Their results suggest that data hiding should be in a domain in which the cover signal components are approximately independent identically distributed and the distortion measure is additive.

Cohen and Lapidoth also treat the interplay between the embedder and the attacker as a game in which the overall net data hiding capacity is the *coding value* of the game.[43] Unlike, O'Sullivan *et al.*, they assume that the encoder and decoder are not cognizant of the attack strategy. The cover signal is assumed to be zero-mean with finite variance, and the embedder and attacker are limited by square error distortions. They show that the achievable data hiding capacity is the same in the case of *blind* and *non-blind* watermarking. The results are elaborated upon and extended in Refs. 44 and 45. In this two-part paper,[44,45] Part I deals with proving the watermark game coding theorems, and Part II, with converse theorems. They define a new measure in the watermarking community called "coding capacity" which is the supremum of all achievable watermarking rates for any sequence of allowable attacks. It is a function of the distortion limit for the embedder, the distortion limit for the attacker and the cover signal distribution†. Assuming the manipulations by the embedder and attacker are limited by energy constraints, the authors show that capacity for blind and non-blind watermarking is the same and achievable if the cover signal is independent, identically distributed and Gaussian.

Cohen and Lapidoth also demonstrate that under the condition that the embedder and attacker modifications are constrained by statistical average power limits (in contrast to energy limits), the watermarking capacity is zero for the watermarking game. Corollaries to their work involve demonstrating that it is suboptimal for an attacker to jam the watermark with a signal independent of the watermarked signal which implies that independent additive noise attacks do not work well, and extensions to the work by Costa.[46]

5. THEMES AND TRENDS

We conclude this review paper with a discussion of some general insights and motifs observed by the authors regarding the area of watermarking theory. Scanning the literature, it is evident that the traditional correlator detector, which can be viewed as the old workhorse for watermark detection, is being left behind for more sophisticated statistical extraction techniques. We see "intelligent" detectors used to undo and characterize attacks for improved reliability. Detector development is an important research area because it is at this stage that the "good-guys" have a practical advantage. During embedding, the embedder has little practical control over the opponent. In fact, some of the optimal codes that have been developed assume, to some extent, a behaviour strategy for the attacker. However, because extraction is performed *after* an attack, the extractor has the last word. By being able to estimate more accurately and undo potential attacks, there is a practical advantage; current detector design research attempts to exploit this opportunity to more effectively achieve optimal performance.

The theoretical work in the area of optimal channel code design for watermarking is an interesting topic because it provides these upper performance bounds. Optimal code constructions for specific attacks lets us know how well we can hope to do which has a strong influence in the way we design/improve practical watermarking algorithms. To obtain tractable closed-form expressions for performance measures, attack assumptions are often conducted. It is unclear, however, at this point how practical these ideas are for real media watermarking applications. New research is beginning to bridge the gap between practical algorithm design and theory.

Recent analysis assumes that the embedder and/or the attacker are aware of the other's strategies. The traditional argument used to validate this assumption is based, in part, on an analogy with Kerckhoff's cryptographic principle that the opponent has knowledge of the particular cryptosystem used, and that protection depends solely on the security of the cryptographic key. However, direct application of this principle can be somewhat misleading for watermarking. The watermarking problem is fundamentally different from cryptography because of its psychological nature and the way in which information is secured. The watermark is protected in *irrelevant* components of a host signal using a steganographic key. In some situations, the key helps prevent an attacker from estimating the hidden information, however, it does not necessarily help prevent its removal. For example, in the case of perfect compression, regardless of knowledge of the key, an attacker could destroy the mark.

†By *distortion limit* we mean the maximum signal change (which can be defined in terms of, say, energy or power depending on the context) that can be induced due to watermark embedding or an attack.

Most techniques surveyed (except Refs. 6 and 24) are applied to image watermarking for proof-of-concept. However, it is uncertain how well the assumptions and models hold up for more diverse multimedia, or how well the algorithms scale for video watermarking. One potential avenue currently being investigated by the authors is application of space-time coding for multimedia and video watermarking. Space-time coding deals with combining both spatial and temporal diversity in traditional communication systems to improve the robustness of the watermark.

Future research would benefit from accounting for implementation aspects of employing the theoretical insights recently developed. For example, practicality of derived approaches in terms of memory and processing power in hardware implementation need to be addressed. In addition, the psychology of perception is not sufficiently considered in the proposed work. Better collaboration with the areas involving information-theoretic perception models will benefit the watermarking community.

REFERENCES

1. D. Kundur, "Implications for high capacity data hiding in the presence of lossy compression," in *Proc. IEEE International Conference on Information Technology: Coding and Computing*, pp. 16–21, March 2000.
2. I. J. Cox, J. Kilian, T. Leighton, and T. Shamoon, "A secure, robust watermark for multimedia," in *Proc. First Int. Workshop on Information Hiding*, R. Anderson, ed., no. 1174 in Lecture Notes in Computer Science, pp. 185–206, May/June 1996.
3. A. Piva, M. Barni, F. Bartolini, and V. Cappellini, "DCT-based watermark recovering without resorting to the uncorrupted original image," in *Proc. IEEE Int. Conference on Image Processing*, vol. 1, pp. 520–523, 1997.
4. T. Kohda, Y. Ookubo, and K. Shinokura, "Digital watermarking through CDMA channels using spread spectrum techniques," in *Proc. IEEE International Symposium on Spread-Spectrum Techniques and Applications*, vol. 2, pp. 671–674, September 2000.
5. P. H. W. Wong, O. C. Au, and J. W. C. Wong, "Image watermarking using spread spectrum technique in log-2 spatio domain," in *Proc. IEEE International Symposium on Circuits and Systems*, vol. 1, pp. 224–272, May 2000.
6. N. Checcacci, M. Barni, F. Bartolini, and S. Basagni, "Robust video watermarking for wireless multimedia communications," in *Proc. IEEE Wireless Communications and Networking Conference*, vol. 3, pp. 1530–1535, 2000.
7. I. J. Cox, M. L. Miller, and A. L. McKellips, "Watermarking as communications with side information," *Proceedings of the IEEE* **87**, pp. 1127–1141, July 1999.
8. M. L. Miller, I. J. Cox, and J. A. Bloom, "Informed embedding: Exploiting image and detector information during watermark insertion," in *Proc. IEEE Int. Conf. on Image Processing*, vol. 3, pp. 1–4, September 2000.
9. S. Voloshynovskiy, F. Deguillaume, and T. Pun, "Optimal adaptive diversity watermarking with channel state estimation," in *Proc. SPIE, Security and Watermarking of Multimedia Contents III*, E. J. Delp and P. W. Wong, eds., vol. 4314, January 2001.
10. S. Voloshynovskiy, F. Deguillaume, and T. Pun, "Content adaptive watermarking based on a stochastic multiresolution image modeling," in *Proc. European Signal Processing Conference*, September 2000.
11. S. Voloshynovskiy, A. Herrigel, N. Baumgaertner, and T. Pun, "A stochastic approach to content adaptive digital image watermarking," *Proc. Third International Workshop on Information Hiding* **1768**, pp. 211–236, September 1999.
12. J. R. Hernández and F. Pérez-González, "Statistical analysis of watermarking schemes for copyright protection of images," *Proceedings of the IEEE* **87**, pp. 1142–1166, July 1999.
13. J. R. Hernández, M. Amado, and F. Pérez-González, "Dct-domain watermarking techniques for still images: Detector performance analysis and a new structure," *IEEE Transactions on Image Processing* **9**, pp. 55–68, January 2000.
14. D. Kundur and D. Hatzinakos, "Diversity and attack characterization for improved robust watermarking," *IEEE Transactions on Signal Processing* **29**, October 2001.
15. J. K. Su and B. Girod, "Power-spectrum condition for energy efficient watermarking," in *Proc. IEEE International Conference on Image Processing*, vol. 1, pp. 301–305, October 1999.
16. J. K. Su and B. Girod, "On the robustness and imperceptibility of digital fingerprints," in *Proc. IEEE International Conference on Multimedia Computing and Systems*, vol. 2, pp. 530–535, June 1999.

17. J. K. Su, "Fundamental performance limits of power-spectrum condition-compliant watermarks," in *Proc. SPIE: Security and Watermarking of Multimedia Contents II*, vol. 3971, pp. 314–325, January 2000.
18. J. R. Hernández, F. Pérez-González, J. M. Rodríguez, and G. Nieto, "Performance analysis of a 2d-multipulse amplitude modulation scheme for data hiding and watermarking of still images," *IEEE Journal on Select Areas in Communications* **16**, pp. 510–524, May 1998.
19. J. P. Linnartz and M. van Dijk, "Analysis of the sensitivity attack against electronic watermarks," *Proc. Second International Workshop on Information Hiding* **1525**, pp. 258–272, April 1998.
20. T. Kalker, J.-P. Linnartz, and M. van Dijk, "Watermark estimation through detector analysis," in *Proc. IEEE Int. Conference in Image Processing*, vol. 1, 1998.
21. J.-P. M. G. Linnartz, A. A. C. Kalker, G. F. G. Depovere, and R. A. Beuker, "A reliability model for the detection of electronic watermarks in digital images," in *Proc. Benelux Symposium on Communication Theory*, pp. 202–209, October 1997.
22. G. Depovere, T. Kalker, and J.-P. Linnartz, "Improved watermark detection reliability using filtering before correlation," in *Proc. IEEE Int. Conference in Image Processing*, vol. 1, 1998.
23. J.-P. Linnartz, T. Kalker, and G. Depovere, "Modelling the false alarm and missed detection rate for electronic watermarks," *Proc. Second International Workshop on Information Hiding* **1525**, pp. 329–343, April 1998.
24. J.-P. Linnartz, T. Kalker, and J. Haitsma, "Detecting electronic watermarks in digital video," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 4, pp. 2071–2074, 1999.
25. J. R. Hernández, F. Pérez-González, and J. M. Rodríguez, "The impact of channel coding on the performance of spatial watermarking for copyright protection," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 5, pp. 2973–2976, 1998.
26. H. V. Poor, *An Introduction to Signal Detection and Estimation*, Springer-Verlag, 2nd ed., 1994.
27. R. J. Clark, "Transform coding of images," New York Academic, 1985.
28. J. Oostveen, T. Kalker, and J.-P. Linnartz, "Optimal detection of multiplicative watermarks," in *Proc. European Signal Processing Conference*, vol. 5, pp. 2973–2976, 2000.
29. J. J. K. Ó Ruanaidh and G. Csurka, "A bayesian approach to spread spectrum watermark detection and secure copyright protection for digital image libraries," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 207–212, June 1999.
30. V. Capellini, M. Bartolini, and M. Barni, eds., *Signal Processing, Special Section on Information Theoretic Aspects of Digital Watermarking*, vol. 81 (6), June 2001.
31. M. Barni, F. Bartolini, A. De Rosa, and A. Piva, "Capacity of full frame DCT image watermarks," *IEEE Transactions on Image Processing* **9**, pp. 1450–1455, August 2000.
32. M. Barni, F. Bartolini, A. De Rosa, and A. Piva, "Capacity of the watermark-channel: How many bits can be hidden within a digital image?," in *Proc. SPIE, Security and Watermarking of Multimedia Contents*, E. J. Delp and P. W. Wong, eds., vol. 3657, pp. 437–448, January 1999.
33. B. Chen and G. W. Wornell, "Achievable performance of digital watermarking systems," in *Proc. IEEE International Conference on Computing and Systems*, vol. 1, pp. 13–18, 1999.
34. B. Chen and G. W. Wornell, "Quantization index modulation: A class of provably good methods of digital watermarking and information embedding," *IEEE Transactions on Information Theory* **47**, pp. 1423–1443, May 2001.
35. D. Kundur, "Energy allocation principles for high capacity data hiding," in *Proc. IEEE International Conference on Image Processing*, vol. 1, pp. 423–426, September 2000.
36. C. Fei, D. Kundur, and R. H. Kwong, "The choice of watermark domain in the presence of compression," in *Proc. IEEE Int. Conf. on Information Technology: Coding and Computing*, pp. 79–84, April 2001.
37. C. Fei, D. Kundur, and R. H. Kwong, "Transform-based hybrid data hiding for improved robustness in the presence of perceptual coding," in *Proc. SPIE, Mathematics of Data/Image Coding, Compression and Encryption IV, with Applications*, M. S. Schmalz, ed., vol. 4475, July 2001.
38. 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.
39. J. A. O'Sullivan, P. Moulin, and J. M. Ettinger, "Information theoretic analysis of steganography," in *Proc. IEEE International Symposium on Information Theory*, p. 297, August 1998.

40. P. Moulin and J. A. O'Sullivan, "Information-theoretic analysis of information hiding," submitted to *IEEE Transactions on Information Theory* , preprint 1999.
41. P. Moulin and J. A. O'Sullivan, "Information-theretic analysis of watermarking," in *Proc. IEEE International Conference on Acustics, Speech and Signal Processing*, vol. 6, pp. 3630–3633, 2000.
42. P. Moulin, M. K. Mihçak, and G.-I. Lin, "An information-theretic model for image watermarking and data hiding," in *Proc. IEEE International Conference on Image Processing*, vol. 3, pp. 667–670, September 2000.
43. A. Cohen and A. Lapidoth, "On the Gaussian watermarking game," in *Proc. IEEE International Symposium on Information Theory*, p. 48, June 2000.
44. A. Cohen and A. Lapidoth, "The Gaussian watermarking game – Part I," submitted to *IEEE Transactions on Information Theory* , preprint 2001.
45. A. Cohen and A. Lapidoth, "The Gaussian watermarking game – Part II," submitted to *IEEE Transactions on Information Theory* , preprint 2001.
46. M. H. M. Costa, "Writing on dirty paper," *IEEE Transactions on Information Theory* **29**, pp. 439–441, January 1983.