

# Digital Watermarking

R. Chandramouli  
Department of ECE  
Stevens Institute of Technology  
Hoboken NJ, 07030

Nasir Memon  
Department of Computer Science  
Polytechnic University  
Brooklyn, NY 11201

Majid Rabbani  
Imaging Research & Advanced Development  
Eastman Kodak Company  
Rochester, NY 14650

## 1 Introduction

The advent of the Internet has resulted in many new opportunities for the creation and delivery of content in digital form. Applications include electronic advertising, realtime video and audio delivery, digital repositories and libraries, and Web publishing. An important issue that arises in these applications is the protection of the rights of all participants. It has been recognized for quite some time that current copyright laws are inadequate for dealing with digital data. This has led to an interest towards developing new copy deterrence and protection mechanisms. One such effort that has been attracting increasing interest is based on *digital watermarking* techniques. Digital watermarking is the process of embedding information into digital multimedia content such that the information (which we call the watermark) can later be extracted or detected for a variety of purposes including copy prevention and control. Digital watermarking has become an active and important area of research, and development and commercialization of watermarking techniques is being deemed essential to help address some of the challenges faced by the rapid proliferation of digital content.

In the rest of this chapter we assume that the content being watermarked is a still image, though most digital watermarking techniques are, in principle, equally applicable to audio and video data. A digital watermark can be *visible* or *invisible*. A visible watermark typically consists of a conspicuously visible message or a company logo indicating the ownership of the image as shown in Figure 1. On the other hand, an invisibly watermarked image appears very similar to the original. The existence of an invisible watermark can only be determined using an appropriate watermark extraction or detection algorithm. In this chapter we restrict our attention to invisible watermarks.

An invisible watermarking technique, in general, consists of an encoding process and a decoding process. A generic watermark encoding process is shown in Figure 2. Here, the watermark insertion step is represented as:

$$X' = E_K(X, W) \tag{1}$$

where  $X$  is the original image,  $W$  is the watermark information being embedded,  $K$  is the user's insertion key, and  $E$  represents the watermark insertion function. We adopt the notation throughout this chapter that for an original image  $X$ , the watermarked variant is represented as  $X'$ . Depending on the way the watermark is inserted, and depending on the nature of the watermarking algorithm,

the detection or extraction method can take on very distinct approaches. One major difference between watermarking techniques is whether or not the watermark detection or extraction step requires the original image. Watermarking techniques that do not require the original image during the extraction process are called *oblivious* (or *public* or *blind*) watermarking techniques. For oblivious watermarking techniques, watermark extraction works as follows:

$$\hat{W} = D_{K'}(\hat{X}') \quad (2)$$

where  $\hat{X}'$  is a possibly corrupted watermarked image,  $K'$  is the extraction key,  $D$  represents the watermark extraction/detection function, and  $\hat{W}$  is the extracted watermark information (see, Figure 3). Oblivious schemes are attractive for many applications where it is not feasible to require the original image to decode a watermark.

Invisible watermarking schemes can also be classified as either *robust* or *fragile*. Robust watermarks are often used to prove ownership claims and so are generally designed to withstand common image processing tasks such as compression, cropping, scaling, filtering, contrast enhancement, printing/scanning, etc., in addition to malicious attacks aimed at removing or forging the watermark.

## 1.1 Applications

Digital Watermarks are potentially useful in many applications, including:

**Ownership assertion.** Watermarks can be used for ownership assertion. To assert ownership of an image, Alice can generate a watermarking signal using a secret private key, and then embed it into the original image. She can then make the watermarked image publicly available. Later, when Bob contends the ownership of an image derived from this public image, Alice can produce the unmarked original image and also demonstrate the presence of her watermark in Bob's image. Since Alice's original image is unavailable to Bob, he cannot do the same. For such a scheme to work, the watermark has to survive image processing operations aimed at malicious removal. In addition, the watermark should be inserted in such a manner that it cannot be forged as Alice would not want to be held accountable for an image that she does not own.

**Fingerprinting.** In applications where multimedia content is electronically distributed over a network, the content owner would like to discourage unauthorized duplication and distribution by embedding a distinct watermark (or a fingerprint) in each copy of the data. If, at a later point in time, unauthorized copies of the data are found, then the origin of the copy can be determined by retrieving the fingerprint. In this application the watermark needs to be invisible and must also be invulnerable to deliberate attempts to forge, remove or invalidate. Furthermore, and unlike the ownership assertion application, the watermark should be resistant to collusion. That is, a group of  $k$  users with the same image but containing different fingerprints, should not be able to collude and invalidate any fingerprint or create a copy without any fingerprint.

**Copy prevention or control.** Watermarks can also be used for copy prevention and control. For example, in a closed system where the multimedia content needs special hardware for copying and/or viewing, a digital watermark can be inserted indicating the number of copies that are permitted. Every time a copy is made the watermark can be modified by the hardware and after a point the hardware would not create further copies of the data. An example of such a system is the Digital Versatile Disc (DVD). In fact, a copy protection mechanism that includes digital

watermarking at its core is currently being considered for standardization and second generation DVD players may well include the ability to read watermarks and act based on their presence or absence.

Another example is in digital cinema, where information can be embedded as a watermark in every frame or a sequence of frames to help investigators locate the scene of the piracy more quickly and point out weaknesses in security in the movie's distribution. The information could include data such as the name of the theater and the date and time of the screening. The technology would be most useful in fighting a form of piracy that's surprisingly common, i.e., when someone uses a camcorder to record the movie as it's shown in a theater, then duplicates it onto optical disks or VHS tapes for distribution.

**Fraud and tamper detection.** When multimedia content is used for legal purposes, medical applications, news reporting, and commercial transactions, it is important to ensure that the content was originated from a specific source and that it had not been changed, manipulated or falsified. This can be achieved by embedding a watermark in the data. Subsequently, when the photo is checked, the watermark is extracted using a unique key associated with the source, and the integrity of the data is verified through the integrity of the extracted watermark. The watermark can also include information from the original image that can aid in undoing any modification and recovering the original. Clearly a watermark used for authentication purposes should not affect the quality of an image and should be resistant to forgeries. Robustness is not critical as removal of the watermark renders the content inauthentic and hence of no value.

**ID card security.** Information in a passport or ID (e.g., passport number, person's name, etc.) can also be included in the person's photo that appears on the ID. By extracting the embedded information and comparing it to the written text, the ID card can be verified. The inclusion of the watermark provides an additional level of security in this application. For example, if the ID card is stolen and the picture is replaced by a forged copy, the failure in extracting the watermark will invalidate the ID card.

The above represent a few example applications where digital watermarks could potentially be of use. In addition there are many other applications in rights management and protection like tracking use of content, binding content to specific players, automatic billing for viewing content, broadcast monitoring etc. From the variety of potential applications exemplified above it is clear that a digital watermarking technique needs to satisfy a number of requirements. Since the specific requirements vary with the application, watermarking techniques need to be designed within the context of the entire system in which they are to be employed. Each application imposes different requirements and would require different types of invisible or visible watermarking schemes or a combination thereof. In the remaining sections of this chapter we describe some general principles and techniques for invisible watermarking. Our aim is to give the reader a better understanding of the basic principles, inherent trade-offs, strengths, and weakness, of digital watermarking. We will focus on image watermarking in our discussions and examples. However as we mentioned earlier, the concepts involved are general in nature and can be applied to other forms of content such as video and audio.

## 1.2 Relationship with Information Hiding and Steganography

In addition to digital watermarking, the general idea of hiding some information in digital content has a wider class of applications that go beyond mere copyright protection and authentication. The techniques involved in such applications are collectively referred to as *information hiding*. For

example, an image printed on a document could be annotated by information that could lead a user to its high resolution version as shown in Figure 4. Metadata provides additional information about an image. Although metadata can also be stored in the file header of a digital image, this approach has many limitations. Usually, when a file is transformed to another format (e.g., from TIFF to JPEG or to bmp), the metadata is lost. Similarly, cropping or any other form of image manipulation destroys the metadata. Finally, the metadata can only be attached to an image as long as the image exists in the digital form and is lost once the image is printed. Information hiding allows the metadata to travel with the image regardless of the file format and image state (digital or analog). Metadata information embedded in an image can serve many purposes. For example, a business can embed the website URL for a specific product in a picture that shows an advertisement for that product. The user holds the magazine photo in front of a low-cost CMOS camera that is integrated into a personal computer, cell phone, or a palm pilot. The data is extracted from the low-quality picture and is used to take the browser to the designated website. Another example is embedding GPS data (about 56 bits) about the capture location of a picture. The key difference between this application and watermarking is the absence of an active adversary. In watermarking applications like copyright protection and authentication, there is an active adversary that would attempt to remove, invalidate or forge watermarks. In information hiding there is no such active adversary as there is no value associated with the act of removing the information hidden in the content. Nevertheless, information hiding techniques need to be robust against accidental distortions. For example, in the application shown in Figure 4, the information embedded in the document image needs to be extracted despite distortions incurred in the print and scan process. But these distortions are just a part of a process and not caused by an active adversary.

Another topic that is related to watermarking is steganography (meaning *covered writing* in Greek), which is the science and art of secret communication. Although steganography has been studied as part of cryptography for many decades, the focus of steganography is secret communication. In fact, the modern formulation of the problem goes by the name of the *prisoner's problem*. Here Alice and Bob are trying to hatch an escape plan while in prison. The problem is that all communication between them is examined by a warden, Wendy, who will place both of them in solitary confinement at the first hint of any suspicious communication. Hence, Alice and Bob must trade seemingly inconspicuous messages that actually contain hidden messages involving the escape plan. There are two versions of the problem that are usually discussed – one where the warden is *passive*, and only observes messages and the other where the warden is *active* and modifies messages in a limited manner to guard against hidden messages. Clearly the most important issue here is that the very presence of a hidden message must be concealed. Whereas in digital watermarking it is not clear that a good watermarking technique should also be steganographic.

### 1.3 Watermarking Issues

The important issues that arise in the study of digital watermarking techniques are:

- *Capacity*: what is the optimum amount of data that can be embedded in a given signal? What is the optimum way to embed and then later extract this information?
- *Robustness*: How do we embed and retrieve data such that it would survive malicious or accidental attempts at removal?
- *Transparency*: How do we embed data such that it does not perceptually degrade the underlying content?

- *Security*: How do we determine that the information embedded has not been tampered, forged or even removed?

Indeed, these questions have been the focus of intense study in the past few years and some remarkable progress has already been made. However, there are still more questions than answers in this rapidly evolving research area. Perhaps a key reason for this is the fact that digital watermarking is inherently a multi-disciplinary topic that builds on developments in diverse subjects. The areas that contribute to the development of digital watermarking include at the very least the following:

- Information and Communication Theory
- Decision and Detection Theory
- Signal Processing
- Cryptography and Cryptographic Protocols

Each of these areas deals with a particular aspect of the digital watermarking problem. Generally speaking, information and communication theoretic methods deal with the data embedding (encoder) side of the problem. For example, information theoretic methods are useful in the computation of the amount of data that can be embedded in a given signal subject to various constraints such as peak power (square of the amplitude) of the embedded data or the embedding induced distortion. The host signal can be treated as a communication channel and various operations such as compression/decompression, filtering etc. can be treated as noise. Using this framework, many results from classical information theory can be and indeed have been successfully applied to compute the data embedding capacity of a signal.

Decision theory is used to analyze data-embedding procedures from the receiver (decoder) side. Given a data-embedding procedure how do we extract the hidden data from the host signal which may have been subjected to intentional or unintentional attacks? The data extraction procedure must be able to guarantee certain amount of reliability. What are the chances that the extracted data is indeed the original embedded data? Even if the data-embedding algorithm is not intelligent or sophisticated, a good data extraction algorithm can offset this effect. In watermarking applications where the embedded data is used for copyright protection, decision theory is used to detect the presence of embedded data. In applications like media bridging, detection theoretic methods are needed to extract the embedded information. Therefore, decision theory plays a very important role in the context of digital watermarking for data extraction and detection. In fact, it is shown that *in the case of using invisible watermarks for resolving rightful ownership, uniqueness problems arise due to the data detection process irrespective of the data embedding process*. Therefore, there is a real and immediate need to develop *reliable, efficient, and robust* detectors for digital watermarking applications.

A variety of signal processing algorithms can be and have been used for digital watermarking. Such algorithms are based on aspects of the human visual system, properties of signal transforms (*e.g.*, Fourier and discrete cosine transform (DCT)), noise characteristics, properties of various signal processing attacks etc. Depending on the nature of the application and the context these methods can be implemented at the encoder, at the decoder, or both. The user has the flexibility to mix and match from different techniques depending on the algorithmic and computational constraints. Although issues such as visual quality, robustness, and real-time constraints can be accommodated, it is still not clear if all the desirable properties for digital watermarking discussed earlier can be achieved by any single algorithm. In most cases these properties have an inherent

trade-off. Therefore, developing signal processing methods to *strike an optimal balance between the competing properties of a digital watermarking algorithm* is necessary.

Cryptographic issues lie at the core of many applications of information hiding but have unfortunately seen little attention. Perhaps this is due to the fact that most work in digital watermarking has been done in the signal processing and communications community whereas cryptographers have focused more on issues like secret communication (covert channels, subliminal channels), and collusion resistant fingerprinting. It is often assumed that simply using appropriate cryptographic primitives like encryption, time-stamps, digital signatures, hash functions, etc. would result in secure information hiding applications. We believe this is far from the truth. In fact, we believe that the design of secure digital watermarking techniques requires an intricate blend of cryptography along with information theory and signal processing.

The rest of this chapter is organized as follows. In Section 2 we describe fragile and semi-fragile watermarking, Section 3 deals with robust watermarks. Communication and information theoretic approaches to watermarking are discussed in Section 4.

## 2 Fragile and Semi-Fragile Watermarks

In the analog world, an image (a photograph) has generally been accepted as a “proof of occurrence” of the depicted event. The advent of digital images and the relative ease with which they can be manipulated, has changed this situation dramatically. Given an image, in digital or analog form, one can no longer be assured of its authenticity. This has led to the need for *image authentication techniques*.

Authentication techniques have been studied in cryptography now for a few decades. They provide a means of ensuring the integrity of a message. So at first sight, the need for image authentication techniques may not seem to pose a problem as many efficient and effective authentication techniques are known from developments in the field of cryptography. Unfortunately, this is far from the truth. Given the large amount of redundancy present in image data, and consequently the large number of different representations of perceptually identical content, the requirement for authentication techniques for images present some unique problems that are not addressed by conventional cryptographic authentication techniques. We list some of these issues below:

- It is desirable in many applications to authenticate the image content, rather than the representation of the content. For example, converting an image from JPEG to GIF is a change in representation. One would like the authenticator to remain valid across different representation as long as the perceptual content has not been changed. Conventional authentication techniques based on cryptographic hash functions, message digests and digital signatures only authenticate the representation.
- When authenticating image content, it is often desirable that the authenticator be embedded in the image itself. One advantage of doing this is that authentication will not require any modifications to the large number of existing representation formats for image content that do not provide any explicit mechanism for including an authentication tag (like the GIF format). However, in our opinion, the most important advantage is that the authentication tag embedded in the image would survive transcoding of the data across different formats, including analog to digital and digital to analog conversions, in a completely transparent manner.
- When authenticating image content, it is desired that one should not only detect the event that the given content has been modified but also detect the exact location where the modification

has taken place.

- Given the highly data intensive nature of image content, any authentication technique has to be computationally efficient to the extent that a simple real-time implementation, both in hardware and software should be possible.

The above issues can be addressed by designing image authentication techniques based on digital watermarks. There are two kinds of watermarking techniques that have been developed for authentication applications - Fragile Watermarking techniques and Semi-Fragile Watermarking techniques. In the rest of this section we describe the general approach taken by each and give some illustrative examples.

## 2.1 Fragile Watermarks

A fragile watermark is designed to indicate and even pin-point any modification made to an image. To illustrate the basic workings of fragile watermarking, we describe a technique recently proposed by Wong and Memon[44]. This technique inserts an invisible watermark  $W$  into an  $m \times n$  image,  $X$ . The original image  $X$  is partitioned into  $k \times l$  blocks, such that  $X_r$  is taken to mean the  $r^{th}$  block of the image; the bi-level watermark  $W$  is partitioned likewise, such that  $W_r$  denotes the  $r^{th}$  block of the watermark. For each image block  $X_r$ , a corresponding block  $\tilde{X}_r$  is formed, identical to  $X_r$  with the exception that the least significant bit of every element in  $\tilde{X}_r$  is set to zero.

For each block  $X_r$ , a cryptographic hash  $H(K, m, n, \tilde{X}_r)$  (such as MD5) is computed, where  $K$  is the user's key. The first  $kl$  bits of the hash output, treated as an  $k \times l$  rectangular array, are XOR'ed with the current watermark block  $W_r$  to form a new binary block  $C_r$ . Each element of  $C_r$  is inserted into the least significant bit of the corresponding element in  $\tilde{X}_r$ , generating the output block  $X'_r$ .

Image authentication is performed by extracting  $C_r$  from each block  $X'_r$  of the watermarked image, and by XOR'ing that array with the cryptographic hash  $H(K, m, n, \tilde{X}_r)$  in a manner similar to above, to produce the extracted watermark block. Changes to the watermarked image result in changes to the corresponding binary watermark region, enabling the technique to be used to localize unauthorized alterations to an image.

The watermarking algorithm can also be extended to a public key version where the private key of a public key algorithm  $K'_A$  is required to insert the watermark. However, the extraction only requires the public key of user  $A$ . More specifically, in the public key version of the algorithm, the MSB's of an image data block  $X_r$  and the image size parameters are hashed, and then the result is encrypted using a public key algorithm. The resulting encrypted block is then XOR'ed with the corresponding binary watermark block  $W_r$  before the combined results are embedded into the LSB of the block. In the extraction step, the same MSB data and the image size parameters are hashed. The LSB of the data block (cipher text) is decrypted using the public key, and then XOR'ed with the hash output to produce the watermark block. Refer to Figure 5 and Figure 6 for public key verification watermark insertion and extraction processes, respectively.

The above technique is just an example of a fragile watermarking technique. There are many more similar techniques proposed in the literature. The main issues that need to be addressed in the design of fragile watermarking techniques are:

- *Locality*: How well does the technique identify the exact pixels that have been modified. The Wong and Memon technique described above, for example, is only able to identify image blocks (at least  $12 \times 12$ ) modified. Any region smaller than this can not be pinpointed as modified.

- *Transparency*: How much degradation in image quality is suffered by insertion of a watermark.
- *Security*: How easy or difficult is it for some one who does not know only the secret key used in the watermarking process to modify an image without modifying the watermark; or by inserting a new but valid watermark.

## 2.2 Semi-Fragile Watermarks

The methods described in the previous subsection authenticate the data that forms the multimedia content, and the authentication process does not treat the data as being distinct from any other data stream. Only the process of inserting the signature into the multimedia content treats the data stream as an object that is to be viewed by a human observer. For example, a watermarking scheme may maintain the overall average image color; or it may insert the watermark in the least significant bit thus discarding the least significant bits of the original data stream and treating them as perceptually irrelevant, or irrelevant to image content.

All multimedia content in current representations have a fair amount of built-in redundancy, that is to say that the data representing the content can be changed without effecting a perceptual change. Further, even perceptual changes to the data may not affect the content. For example, when dealing with images, one can brighten an image, compress it in a lossy fashion, or change contrast settings. The changes caused by these operations could well be perceptible, even desirable, but the image content is not considered changed. Objects in the image are in the same positions as well as setting and are still recognizable. It is highly desirable that authentication of multimedia documents take this into account - that is, there be a set of ‘allowed’ operations, and ‘image content’; it is with respect to allowing the first and retaining the second that any authentication should be performed for it to be genuinely useful.

There have been a number of recent attempts at techniques which address authentication of ‘image content’, and not of only image data. One approach is to use feature points in defining image content that is robust to image compression. An image authentication scheme for image content would then be one which used cryptographic schemes like digital signatures to authenticate these feature points. Typical feature points include, for example, edge maps[2], local maximas and minimas and low pass wavelet coefficients[24]. The problems with these methods is that it is hard to define image content in terms of a few features; for example edge maps do not sufficiently define image content as it may be possible to have two images with fairly different content (the face of one person replaced by that of another) but with identical edge maps. Image content remains an extremely ill-defined quantity despite the attempts of the vision and compression communities to nail it down.

Another interesting approach for authenticating image content is to compute an image digest (or hash or fingerprint) of the image and then encrypt the digest with a secret key. For public key verification of the image, the secret key is the user’s private key and hence the verification can then be done by anyone with the user’s public key, much like digital signatures. It should be noted that the image digest that is computed is much smaller than the image itself and can be embedded into the image using a robust watermarking technique. Furthermore, the image digest has the property that as long as the image content has not changed the digest that is computed from the image remains the same. Clearly constructing such an image digest function is a difficult problem. Nevertheless, there have been a few such functions proposed in the literature and image authentication schemes based on them have been devised. Perhaps the most widely cited image digest function/authentication scheme is SARI, proposed by Lin and Chang [26] . The SARI authentication scheme contains an image digest function that generates hash bits that are invariant

to JPEG compression. That is, the hash bits do not change if the image is JPEG compressed but do change for any other significant or malicious operation.

The image digest component of SARI is based on the invariance of the relationship between selected DCT coefficients in two given image blocks. It can be proven that this relationship is maintained even after JPEG compression using the same quantization matrix for the whole image. Since the image digest is based on this feature, SARI can distinguish between JPEG compression and other malicious operations that modify image content. More specifically, in SARI, the image to be authenticated is first transformed to the DCT domain. The DCT blocks are grouped into non-overlapping sets  $P_p$  and  $P_q$  as defined below:

$$P_p = \{P_1, P_2, P_3, \dots, P_{\frac{N}{2}}\}$$

$$P_q = \{Q_1, Q_2, Q_3, \dots, Q_{\frac{N}{2}}\}$$

where  $N$  is the total number of DCT blocks in the input image. An arbitrary mapping function,  $Z$ , is defined between these two sets satisfying the following criteria  $P_p = Z(K, P_q)$ ,  $P_p \cap P_q = \emptyset$  and  $P_p \cup P_q = P$  where  $P$  is the set of all DCT blocks of the input image. The mapping function,  $Z$ , is central to the security of SARI and is not publicized. In fact, it is based on a secret key  $K$ . The mapping effectively partitions image blocks into pairs. Then for each block pair, a number of DCT coefficients are selected. Feature code or hash bits are then generated by comparing the corresponding coefficients in the paired block. For example, in the block pair  $(P_m, P_n)$  if the DC coefficient in block  $P_m$  is greater than the DC coefficient in block  $P_n$ , then the hash bit generated is '1'. Otherwise, a '0' is generated.

It is clear that a hash bit serves to preserve the relationship between the selected DCT coefficients in a given block pair. The hash bits generated for each block are concatenated to form the digest of the input image. This digest can then be either embedded into the image itself or appended as a tag. The authentication procedure at the receiving end involves the extraction of embedded digest. The digest for the received image is generated as at the encoder and compared with the extracted and decrypted digest. Since relationships between selected DCT coefficients are maintained even after JPEG compression, this authentication system can distinguish JPEG compression from other malicious manipulations on the authenticated image. However, it was recently shown that if a system uses the same secret key  $K$  and hence the same mapping function  $Z$  to form block pairs for all the images authenticated by it, an attacker with access to a sufficient number of images authenticated by this system can produce arbitrary fake images [34].

SARI is limited, by design to authenticate only after compression. Although compression is the most common operation that may be carried out on an image, certain applications may require authentication to be performed after other simple image processing operations like sharpening, deblurring etc. Again many techniques have been proposed but perhaps the best known one is by Fridrich [16]. In this technique,  $N$  random matrices are generated with entries uniformly distributed in  $[0,1]$ , using a secret key. Then, a low-pass filter is repeatedly applied to each of these random matrices to obtain  $N$  random smooth patterns as shown in Figure 8. These are then made DC free by subtracting their respective means to obtain  $P_i$  where  $i = 1, \dots, N$ . Then image block,  $B$ , is projected on to each of these random smooth patterns. If a projection is greater than zero then the hash bit generated is a '1' otherwise a '0' is generated. In this way a  $N$  bit hash are generated for image authentication.

Since the patterns  $P_i$  have zero mean, the projections do not depend on the mean gray value of the block and only depend on the variations within the block itself. The robustness of this bit extraction technique was tested on real imagery and it was shown that it can reliably extract over 48 correct bits (out of 50 bits) from a small  $64 \times 64$  image for the following image processing operations:

15% quality JPEG compression (as in PaintShop Pro), additive uniform noise with amplitude of 30 gray levels, 50% contrast adjustment, 25% brightness adjustment, dithering to 8 colors, multiple applications of sharpening, blurring, median, and mosaic filtering, histogram equalization and stretching, edge enhancement, and gamma correction in the range 0.7-1.5. However, operations like embossing and geometrical modifications, such as rotation, shift, and change of scale, lead to a failure to extract the correct bits.

In summary, image content authentication using a visual hash function and then embedding this hash using a robust watermark is a promising area and will see many developments in the coming years. This is a difficult problem and we doubt if there will ever be a completely satisfactory solution. The main reason for this being that there is no clear definition of image content and small changes to an image could potentially lead to different content.

### 3 Robust Watermarks

Unlike fragile watermarks, robust watermarks are resilient to intentional or un-intentional attacks or signal processing operations. Ideally, it must withstand attempts to destroy or remove it. Some of the desirable properties of a good, robust watermark include the following :

- Perceptual transparency: Robustness must not be achieved at the expense of perceptible degradation to the watermarked data. For example, a high energy watermark can withstand many signal processing attacks; however, even in the absence of any attacks this can cause significant loss in the visual quality of the watermarked image.
- Higher pay load: A robust watermark must be able to reliably carry higher number of information bits even in the presence of attacks.
- Resilience to common signal processing operations such as compression, linear and non-linear filtering, additive random noise, digital to analog conversion etc.
- Resilience to geometric attacks such as translation, rotation, cropping, and scaling.
- Robustness against collusion attacks where multiple of copies of the watermarked data can be used to create a valid watermark.
- Computational simplicity: Consideration for computational complexity is important while designing robust watermarks. If a watermarking algorithm is robust but computationally very intensive during encoding or decoding then its usefulness in real-life may be limited.

Of course, the above features do not come for free. There are a number of tradeoffs. Three major tradeoffs in robust watermarking and the applications that are impacted by each of these tradeoff factors are shown in Figure 9.

It is easily understood that placing a watermark in perceptually insignificant components of an image causes imperceptible distortions to the watermarked image. But, we observe that, such watermarking techniques are not robust against intentional or unintentional attacks. For example, if the watermarked image is lossy compressed then the perceptually insignificant components are discarded by the compression algorithm. Therefore, for a watermark to be robust, it must be placed in the perceptually significant components of an image even though we run a risk of causing perceptible distortions. This gives rise to two important questions: (a) what are the perceptually significant components of a signal, (b) how can the perceptual degradation due to robust watermarking be minimized ? The answer to the first question depends on the type of media—audio,

image, or video. For example, certain spatial frequencies and some spatial characteristics such as edges in an image are perceptually significant. Therefore, choosing these components as carriers of a watermark will add robustness against operations such as lossy compression.

There are a multitude of ways in which a watermark can be inserted into the perceptually significant components. But, care must be taken to *shape* the watermark to match the characteristics of the carrier components. A common technique that is used in most robust watermarking algorithms is the adaptation of the watermark energy to suit the characteristics of the carrier. This is usually done based on certain local statistics of the original image such that the watermark is not visually perceptible.

There have been a number of robust watermarking techniques developed in the past few years. Some of these are in the spatial domain and some in the frequency domain. Some are additive watermarks and some use a quantize and replace strategy. Some are linear and some are non-linear. The earliest robust spatial domain techniques were perhaps the MIT patchwork algorithm [1] and another one by Digimarc [10]. One of the first and perhaps still the most cited frequency domain technique was proposed by Cox et. al. [12]. Some early perceptual watermarking techniques using linear transforms in the transform domain were proposed in [43]. Finally some recent and remarkably robust techniques were proposed by Kodak in [14, 22, 13, 20, 21]. Instead of describing these different algorithms independently, we instead choose to describe Kodak’s technique in detail as it clearly identifies the different elements that are needed in a robust watermarking technique.

### Kodak’s Watermarking Technique

An example of a spatial watermarking technique is one based on phase dispersion that has been developed by Kodak [14, 22, 13, 20, 21]. The Kodak method is noteworthy for several reasons. The first is that it can be used to embed either a grayscale iconic image or binary data. Iconic images include trademarks, corporate logos or other arbitrary small images and an example is shown in Figure 10. The second is that the technique can determine cropping coordinates without the need for a separate calibration signal. Furthermore, the strategy that is used to detect rotation and scale can be applied to other watermarking methods in which the watermark is inserted as a periodic pattern in the image domain. Finally, the Kodak algorithm has scored a reported score of 0.98 using StirMark 3.0 [21]. The following is a brief description of the technique. For brevity, only the embedding of binary data is considered.

The binary digits are represented by positive and negative delta functions (corresponding to ones and zeros) that are placed in unique locations within a message image  $M$ . These locations are specified by a predefined *message template*  $T$ , an example of which is shown in Figure 10. The size of the message template is typically only a portion of the original image size (e.g.,  $64 \times 64$ , or  $128 \times 128$ ). Next, a carrier image  $\tilde{C}$ , which is of the same size as the message image, is generated using a secret key. The carrier image is usually constructed in the Fourier domain by assigning a uniform amplitude and a random phase (produced by a random number generator initialized by the secret key) to each spatial frequency location. The carrier image is convolved with the message image to produce a dispersed message image, which is then added to the original image. Because the message image is typically smaller than the original image, the original image is partitioned into contiguous non-overlapping rectangular blocks,  $X_r$ , which are the same size as the message image. The message embedding process creates a block of the watermarked image,  $X'_r(x, y)$ , according to the following relationship:

$$X'_r(x, y) = \alpha(M(x, y) * \tilde{C}(x, y)) + X_r(x, y) \tag{3}$$

where the symbol  $*$ , represents cyclic convolution and  $\alpha$  is an arbitrary constant chosen to make

the embedded message simultaneously invisible and robust to common processing. This process is repeated for every block in the original image as depicted in Figure 11. From Eq. (3) it is clear that there are no restrictions on the message image, and its pixel values can be either binary or multilevel.

The basic extraction process is straightforward and consists of correlating a watermarked image block with the same carrier image used to embed the message. The extracted message image  $\hat{M}'(x, y)$  is given by:

$$\hat{M}'(x, y) = X_r'(x, y) * \tilde{C}(x, y) = \alpha(M(x, y) * \tilde{C}(x, y)) \otimes \tilde{C}(x, y) + X_r(x, y) \otimes \tilde{C}(x, y) \quad (4)$$

where the symbol  $\otimes$  represents cyclic correlation. The correlation of the carrier with itself can be represented by a point spread function  $p(x, y) = \tilde{C}(x, y) \otimes \tilde{C}(x, y)$ , and since the operations of convolution and correlation commute, Eq. (4) reduces to:

$$\hat{M}'(x, y) = \alpha M(x, y) * p(x, y) + X_r(x, y) \otimes \tilde{C}(x, y) \quad (5)$$

The extracted message is a linearly degraded version of the original message plus a low amplitude noise term resulting from the cross correlation of the original image with the carrier. The original message can be recovered by using any conventional restoration (deblurring) technique such as Wiener filtering. However, for an ideal carrier,  $p(x, y)$  is a delta function, and the watermark extraction process results in a scaled version of the message image plus low amplitude noise. To improve the signal to noise ratio of the extraction process, the watermarked image blocks are aligned and summed prior to the extraction process as shown in Figure 12. The summation of the blocks reinforces the watermark component (because it is the same in each block), while the noise component is reduced because the image content typically varies from block to block. In order to create a system that is robust to cropping, rotation, scaling, and other common image processing tasks such as sharpening, blurring, compression, etc., many factors need to be considered in the design of the carrier and the message template.

In general, the design of the carrier requires consideration of the visual transparency of the embedded message, the extracted signal quality, and the robustness to image processing operations. For visual transparency, most of the carrier energy should be concentrated in the higher spatial frequencies since the contrast sensitivity function (CSF) of the human visual system falls off rapidly at higher frequencies. However, to improve the extracted signal quality, the autocorrelation function of the carrier,  $p(x, y)$ , should be as close as possible to a delta function, which implies a flat spectrum. In addition, it is desirable to spread out the carrier energy over all frequencies to improve robustness to both friendly and malicious attacks. This is because the power spectrum of typical imagery falls off with spatial frequency and concentration of the carrier energy in high frequencies would create little frequency overlap between the image and the embedded watermark, rendering the watermark vulnerable to removal by simple low-pass filtering. The actual design of the carrier is a balancing act between these concerns.

The design of an optimal message template is guided by two requirements. The first is to maximize the quality of the extracted signal, which is achieved by placing the message locations maximally apart. The second is that the embedded message must be recoverable from a cropped version of the watermarked image. Consider a case where the watermarked image has been cropped such that the watermark tiles in the cropped image are displaced with respect to the tiles in the original image. It can be shown that the extracted message from the cropped image is a cyclically shifted version of the extracted message from the uncropped image. Since the message template is known, the amount of the shift can be unambiguously determined by insuring that all the cyclic shifts of the message template are unique. This can be accomplished by creating a message template

that has an autocorrelation equal to a delta function. Although in practice it is impossible for the autocorrelation of the message template to be an ideal delta function, optimization techniques such as simulated annealing can be used to design a message template with maximum separation and minimum sidelobes.

The ability to handle rotation and scaling is a fundamental requirement of robust data embedding techniques. Almost all applications that involve printing and scanning will result in some degree of scaling and rotation. Many algorithms rely on an additional calibration signal to correct for rotation and scaling, which taxes the information capacity of the embedding system. Instead, the Kodak approach uses the autocorrelation of the watermarked image to determine the rotation and scale parameters, which does not require a separate calibration signal. This method also can be applied to any embedding technique where the embedded image is periodically repeated in tiles. It can also be implemented over local regions to correct for low order geometric warps.

To see how this method is applied, consider the autocorrelation function of a watermarked image that has not been rotated or scaled. At zero displacement, there is a large peak due to the image correlation with itself. However, since the embedded message pattern is repeated at each tile, lower magnitude correlation peaks are also expected at regularly spaced horizontal and vertical intervals equal to the tile dimension. Rotation and scaling affect the relative position of these secondary peaks in exactly the same way that they affect the image. By properly detecting these peaks, the exact amount of the rotation and scale can be determined. An example is shown in Figure 13. Not surprisingly, the energy of the original image is much larger than that of the embedded message, and the autocorrelation of the original image can mask the detection of the periodic peaks. To minimize this problem, the watermarked image needs to be processed, prior to the computation of the autocorrelation function. Examples of such pre-processing include removal of the local mean by a spatially adaptive technique or simple high-pass filtering. In addition, the resulting autocorrelation function is high-pass filtered to amplify the peak values.

## 4 Communication and Information Theoretic Aspects

Communication and information theoretic approaches focus mainly on the theoretical analysis of watermarking systems. They deal with abstract mathematical models for watermark encoding, attacks, and decoding. These models enable the study of watermarks at a high level without resorting to any specific application (such as image watermarking etc.). Therefore, the results obtained using these techniques are potentially applicable to a wide variety of application scenarios by suitably mapping the application to a communication or information theoretic model. The rich set of mathematical models primarily based on the theory of probability and stochastic processes allows a rigorous study of watermarking techniques; however, a common complaint from practitioners suggests that some of these popular mathematical theories are not completely valid in practice. Therefore, we observe that, studying watermarks based on communication and information theory is an on-going process where theories are proposed and refined based on feedback from engineering applications of watermarks.

In this section we describe some communication and information theoretic aspects of digital watermarking. We first describe the similarities and differences between classical communication and current watermarking systems. Once this is established it becomes easier to adapt the theory of communications to watermarking and make theoretical predictions about the performance of a watermarking system. Following this discussion we describe some information theoretic models applied to watermarking.

## 4.1 Watermarking as Communication

It is quite common and popular to adapt techniques from standard communication theory to study and improve watermarking algorithms [11] using models similar to the ones shown in Figure 14 and Figure 15. Figure 14 shows how the information bits are first encoded (to suit the modulation type, error control etc.) followed by modulating a carrier signal that carries this information across a noisy channel. At the decoder side, this carrier is demodulated and then the information (possibly corrupted due to channel noise) is decoded. In a digital watermarking system as seen in Figure 15 we see that the modulator in Figure 14 is replaced by the watermark embedder that places the watermark in the media content. Distortions to the watermarked media is induced by known or unknown attacks or signal processing operations such as compression, decompression, cropping, scaling etc. The embedded watermark is finally retrieved by the watermarked decoder or detector. One major difference between the two models can be seen in the encoder side. While, in communication systems, the encoding is done in order to protect the information bits from channel distortion, in watermarking, emphasis is usually placed on techniques that minimize perceptual distortions to the watermarked content.

Some analogies between the traditional communication system and the watermarking system are summarized in Table 1. From this table we note that the theory and algorithms developed for

Table 1: Analogies between communication and watermarking system.

Communication System	Watermarking System
Information	Watermark
Communication channel	Host signal (such as image, video etc.)
Power constraint on transmitted signal due to physical limitations	Power constraint on watermark due to audio/visual quality limitations
Interference	Host signal and watermark attacks
Side information at transmitter and/or receiver	Knowledge of host signal, watermarking parameters such as key etc. at the encoder and/or decoder.
Channel capacity	Watermarking capacity

the study of digital communication systems may be directly applicable to study some aspects of watermarking. Note that these two systems have common constraints such as power and reliability while they differ in constraints such as perceptual constraints (applicable only to watermarking).

## 4.2 Information Theoretic Analysis

Information theoretic methods have been applied to information storage and transmission with great success [36]. Here, messages and channels are modelled probabilistically and their properties are studied analytically. A great amount of effort in the past five decades has produced many interesting results regarding the capacity of various channels, *i.e.*, the maximum amount of information that can be transmitted through a channel such that decoding this information with arbitrarily small probability of error is possible. Using the analogy between communication and watermarking channels, it is possible to compute fundamental information carrying capacity limits of watermarking channels using information theoretic analysis. In this context, the following two

important questions arise:

- What is the maximum length (in bits) of a watermark message that can be embedded and distinguished reliably in a host signal?
- How do we design watermarking algorithms that can effectively achieve this maximum?

Answers to these questions can be found at least under certain assumptions [3],[4], [5],[6],[7],[8], [29],[32],[33],and [35]. We usually begin by assuming probability models for the watermark signal, host signal, and the random watermark key. A distortion constraint is then placed on the watermark encoder. This constraint is used to model and control the perceptual distortion induced due to watermark insertion. For example, in image or video watermarking, the distortion metric could be based on human visual perceptual criteria. Based on the application, the watermark encoder can use a suitable distortion metric and a value for this metric that must be met during encoding. A watermark attacker has a similar distortion constraint so that the attack does not result in a completely corrupted watermarked signal making it useless for all concerned parties. The information that is known to the encoder, attacker, and the decoder is incorporated into the mathematical model through joint probability distributions. Then, the *watermarking capacity* is given by the maximum rate of reliable embedding of the watermark over any possible watermarking strategy and any attack that satisfies the specified constraints. This problem can also be formulated as a stochastic game where the players are the watermark encoder and the attacker [9]. The common payoff function of this game is the mutual information between the random variables representing the input and the received watermark.

We now discuss the details of the mathematical formulation described above. Let a watermark (or message)  $W \in \mathcal{W}$  be communicated to the decoder. This is embedded in a length- $N$  sequence  $X^N = (X_1, X_2, \dots, X_N)$  representing the host signal. Let the watermark key known both to the encoder and the decoder be  $K^N = (K_1, K_2, \dots, K_N)$ . Then, using  $W$ ,  $X^N$ , and  $K^N$  a watermarked signal  $X'^N = (X'_1, X'_2, \dots, X'_N)$  is obtained by the encoder. For instance, in transform based image watermarking, each  $X_i$  could represent a block of  $8 \times 8$  discrete cosine transform coefficients,  $W^N$  could be the spread spectrum watermark [12], and  $K^N$  could be locations of the transform coefficients where the watermark is embedded. Therefore,  $N=4096$  for a  $512 \times 512$  image. Usually, it is assumed that the elements of  $X^N$  are independent and identically distributed (i.i.d.) random variables with probability mass function  $p(x)$ ,  $x \in \mathcal{X}$ . Similarly, the elements of  $K^N$  are i.i.d. with probability mass function  $p(k)$ ,  $k \in \mathcal{K}$ . If  $X$  and  $K$  denote generic random variables in the random vectors  $X^N$  and  $K^N$ , respectively, then any dependence between  $X$  and  $K$  are modelled by the joint probability mass function,  $p(x, k)$ . Usually,  $W$  is assumed to be independent of  $(X, K)$ . Then a length- $N$  watermarking code with distortion  $D_1$  is a triple  $(\mathcal{W}, f_N, \phi_N)$ , where,  $\mathcal{W}$  is a set of messages with uniformly distributed elements,  $f_N$  is the encoder mapping, and  $\phi_N$  is the decoder mapping that satisfy the following [29]:

- The encoder mapping  $x'^N = f_N(x^N, w, k^N) \in \mathcal{X}^N$  is such that the expected value of the distortion,  $E[d^N(X^N, X'^N)] \leq D_1$ .
- The decoder mapping is given by  $\hat{w} = \phi_N(y^N, k^N) \in \mathcal{W}$  where  $y^N$  is the received watermarked signal.

The attack channel is modelled as a sequence of conditional probability mass functions,  $A^N(y^N|x^N)$  such that  $E[d^N(X^N, Y^N)] \leq D_2$ . Throughout it is assumed that  $d^N(x^N, y^N) = \frac{1}{N} \sum_{j=1}^N d(x_j, y_j)$  where  $d$  is a bounded, non-negative, real-valued distortion function. A watermarking rate  $R = \frac{1}{N} \log |\mathcal{W}|$  is said to be achievable for  $(D_1, D_2)$  if there exists a sequence of watermarking codes

$(\mathcal{W}, f_N, \phi_N)$  subject to distortion  $D_1$  with respective rates  $R_N > R$  such that the probability of error  $P_e = \frac{1}{|\mathcal{W}|} \sum_{w \in \mathcal{W}} Pr(\hat{w} \neq w | W = w) \rightarrow 0$  as  $N \rightarrow \infty$  for any attack subject to  $D_2$ . The watermarking capacity  $C(D_1, D_2)$  is then defined as maximum (or supremum, in general) of all achievable rates for given  $D_1$  and  $D_2$ . This information theoretic framework has been successfully used to compute the watermarking capacity of a wide variety of channels. We discuss a few of them next.

When  $N = 1$  in the information theoretic model we obtain a single letter channel. Consider the single letter, discrete time, additive channel model shown in Figure 16. In this model, the message  $W$  is corrupted by additive noise  $J$ . Suppose  $E(W) = E(J) = 0$  then the watermark power is given by  $E(W^2) = \sigma_W^2$  and channel noise power is  $E(J^2) = \sigma_J^2$ . If  $W$  and  $J$  are Gaussian distributed then it can be shown that the watermarking capacity is given by  $1/2 \ln \left( 1 + \frac{\sigma_W^2}{\sigma_J^2} \right)$  [35]. For the Gaussian channel case a surprising result has also been found recently [29]. Let  $\mathcal{W} = \mathfrak{R}$  be the space of the watermarked signal and  $d(w, y) = (w - y)^2$  be the squared-error distortion measure. If  $X \sim \text{Gaussian}(0, \sigma_x^2)$  then the capacity of the blind and non-blind watermarking systems are equal! This means that irrespective of whether the original signal is available at the decoder or not the watermarking rate remains the same.

The watermarking capacity when the host signal undergoes specific kinds of processing/attacks that can be modeled using well-known probability distributions have received considerable attention. Also, a popular assumption is that the type of attack the watermarked signal undergoes is completely known at the receiver and is usually modeled as additive noise. But, in reality, an attack is not guaranteed to be known at the receiver, and, it need not be additive only; *e.g.* scaling and rotation attacks are not additive. Therefore, a more general mathematical model as shown in Figure 17 is required to improve the capacity estimates for many non-additive attack scenarios [4]. We see in Figure 17 that a random multiplicative component is also introduced to model an attack.

Using the model seen in Figure 17 where  $G_d$  and  $G_r$  respectively denote the deterministic and random components of the multiplicative channel noise attack it has been shown that [4] a traditional additive channel model such as the one shown in Figure 16 tends to either over or under-estimate the watermarking capacity depending on the type of attack. A precise estimate for the loss in the capacity due to the uncertainty about the channel attack at the decoder can be computed using this model. Extensions of this result to multiple watermarks in a host signal show that, in order to improve the capacity, a specific watermark decoder has to cancel the effect of the interfering watermarks rather than treating them as known or unknown interference. It has also been observed that [4] an unbounded increase in watermark energy does not necessarily produce unbounded capacity. These results give us intuitive ideas to optimize watermarking systems for optimum capacity.

Information theoretic watermarking capacity computations do not tell us how to approach this capacity effectively. To address this important problem new set of techniques are required. Approaches such as quantization index modulation (QIM) [7] address some of these issues. QIM deals with the characterization of the inherent trade-offs among embedding rate, embedding-induced degradation, and robustness of embedding methods. Here, the watermark embedding function is viewed as an ensemble of functions indexed by  $w$  that satisfies the following property:

$$x \approx x' \quad \forall w. \quad (6)$$

It is clear that robustness can be achieved if the ranges of these function are sufficiently *separated* from each other. If not, identifying the embedded message uniquely even in the absence of any attacks will not be possible. Eq. (6) and the non-overlapping ranges of the embedding functions suggest that the range of the embedding functions must cover the range space of  $x'$  and the functions

be discontinuous. QIM embeds information by first modulating an index or a sequence of indices with the embedding information and then quantizing the host signal with an associated quantizer or a sequence of quantizers. We explain this with an example. Consider the case where one bit is to be embedded, i.e.,  $w \in \{1, 2\}$ . Thus two quantizers are required with their corresponding reconstruction points in  $\mathfrak{R}^N$  well separated in order to inherit robustness against attacks. If  $w = 1$  the host signal is quantized with the first quantizer if not the second quantizer is used. Therefore we see that the quantizer reconstruction points also act as constellation points that carry information. Thus QIM design can be interpreted as the joint design of an ensemble of source codes and channel codes. The number of quantizers determine the embedding rate. It is observed that QIM structures are optimal for memoryless watermark channels when energy constraints are placed on the encoder. As we can see, a fundamental principle behind QIM is the attempt to optimally trade-off embedding rate for robustness.

As discussed in previous sections, many popular watermarking schemes are based on signal transforms such as the discrete cosine transform and wavelet transform. The transform coefficients play the role of carriers of watermarks. Naturally, different transforms possess widely varying characteristics. Therefore a natural question to ask is — what is the effect of the choice of transforms on the watermarking capacity? Note that, good energy compacting transforms such as the discrete cosine transform produce transform coefficients with unbalanced statistical variances. This property is observed to enhance the watermarking capacity in some cases [32]. Results such as these could help us in designing high capacity watermarking techniques that are compatible with transform based data compression standards such as JPEG2000 and MPEG-4.

To summarize, we have seen that communication and information theoretic approaches provide us with valuable mathematical tools to analyze watermarking techniques. They make it possible to predict or estimate the theoretical performance of a watermarking algorithm independent of the underlying application. But, the practical utility of these models and analysis has been questioned by practicing engineers. Therefore, it is important that the developers of mathematical theories for watermarking and real-life system developers must interact with each other through a constructive feedback mechanism to improve the state-of-the-art in digital watermarking technologies.

## 5 Conclusions

Digital watermarking is a rapidly evolving area of research and development. We only discussed the key problems in this area and presented some known solutions in this chapter. One key research problem that we still face today is the development of truly robust, transparent and secure watermarking technique for different digital media including images, video and audio. Another key problem is the development of semi-fragile authentication techniques. The solution to these problem will require application of known results and development of new results in the fields of information and coding theory, adaptive signal processing, game theory, statistical decision theory, and cryptography. Although a lot of progress has already been made, there still remain many open issues that need attention before this area becomes mature. This chapter has only provided a snapshot of the current state-of-the-art. For details the reader is referred to the survey articles [23, 28, 19, 18, 15, 17, 25, 39, 45, 31, 38, 41] that deal with various important topics and techniques in digital watermarking. We hope these references will be useful both to a newcomer and an advanced researcher.

## References

- [1] W. Bender D. Gruhl N. Morimoto and A. Lu. Techniques for data hiding. *IBM Systems Journal*, 35(3-4):313–336, 1996.
- [2] S. Bhattacharjee, “Compression Tolerant Image Authentication”, *Proceedings, Int. Conf. Image Proc.*, Chicago, Oct. 1998.
- [3] C. Cachin. An information-theoretic model for steganography. *Proc. of 2nd Workshop on information hiding*, 1998.
- [4] R. Chandramouli. Data hiding capacity in the presence of an imperfectly known channel. *Proc. SPIE Security and Watermarking of Multimedia Contents III*, 2001.
- [5] R. Chandramouli. Watermarking capacity in the presence of multiple watermarks and partially known channel. *Proc. of SPIE Multimedia Systems and Applications IV*, 4518, Aug. 2001.
- [6] B. Chen and G.W. Wornell. Digital watermarking and information embedding using dither modulation. *IEEE Second Workshop on Multimedia Signal Processing*, pages 273–278, 1998.
- [7] B. Chen and G.W. Wornell. Achievable performance of digital watermarking systems. *IEEE International Conference on Multimedia Computing and Systems*, 1:13–18, 1999.
- [8] B. Chen and G.W. Wornell. An information-theoretic approach to the design of robust digital watermarking systems. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 4:2061–2064, 1999.
- [9] A. Cohen and A. Lapidoth. On the gaussian watermarking game. *Proc. Intl. Symposium on Information Theory*, page 48, June 2000.
- [10] Digimarc Corporation. <http://www.digimarc.com>.
- [11] I. J. Cox, M. L. Miller, and A. L. McKellips. Watermarking as communications with side information. *Proceedings of the IEEE*, 87:1127–1141, July 1999.
- [12] I.J. Cox, J. Kilian, T.Leighton, and T. Shamoon. Secure spread spectrum watermarking for multimedia. *IEEE Transactions on Image Processing*, 6:12, December, 1997, pp. 1673-1687.
- [13] S. J. Daly. Method and apparatus for hiding one image or pattern within another. U.S. Patent 5,905,819, 1999.
- [14] S. J. Daly, J. R. Squilla, M. Denber, C. W. Honsinger, and J. Hamilton. Method for embedding digital information in an image. U.S. Patent 5,859,920, 1999.
- [15] J. Dittmann, A. Behr, M. Stabenau, P. Schmitt, J. Schwenk, and J. Ueberbe. Combining digital watermarks and collusion secure fingerprints for digital images. *SPIE Security and Watermarking of Multimedia Contents*, pages 171–182, 1999.
- [16] J. Fridrich, “Image Watermarking for Tamper Detection”, *Proceedings, Int. Conf. Image Proc.*, Chicago, Oct. 1998.
- [17] J. Fridrich and M. Goljan. Comparing robustness of watermarking techniques. *SPIE Security and Watermarking of Multimedia Contents*, pages 214–225, 1999.

- [18] F. Hartung, J. K. Su, and B. Girod. Spread spectrum watermarking: Malicious attacks and counterattack. *SPIE Security and Watermarking of Multimedia Contents*, pages 147–158, 1999.
- [19] M. Holliman, N. Memon, and M. M. Yeung. Watermark estimation through local pixel estimation. *SPIE Security and Watermarking of Multimedia Contents*, pages 134–146, Jan. 1999.
- [20] C. W. Honsinger and S. J. Daly. Method for detecting rotation and magnification in images. U.S. Patent 5,835,639, 1998.
- [21] C. W. Honsinger and Majid Rabbani. Data embedding using phase dispersion. *International Conference on Information Technology: Coding and Computing*, March 2000.
- [22] C. W. Honsinger and Majid Rabbani. Method for generating an improved carrier for the data embedding problem. U.S. Patent 6,044,156, 2000.
- [23] P. Jessop. The business case for audio watermarking. *International Conference on Acoustics, Speech and Signal Processing*, 80:2077–2080.
- [24] D. Kundur and D. Hatzinakos. Digital Watermarking for Telltale Tamper-Proofing and Authentication. *Proceedings of the IEEE*, Special Issue on Identification and Protection of Multimedia Information, vol. 87, no. 7, pp. 1167-1180, July 1999.
- [25] M. Kutter and F. A. P. Petitcolas. A fair benchmark for image watermarking systems. *SPIE Security and Watermarking of Multimedia Contents*, pages 226–239, 1999.
- [26] C.Y.Lin and S.F.Chang. A Robust Image Authentication Method Distinguishing JPEG Compression from Malicious Manipulation, SPIE Storage and retrieval of Image/Video Databases, San Jose, January 1998.
- [27] N. Memon M. Holliman and M. Yeung. On the need for image dependent keys in watermarking. *Proceedings of the Second Workshop on Multimedia*, March 1999.
- [28] F. Mintzer and G. W. Braudaway. If one watermark is good, are more better? *International Conference on Acoustics, Speech and Signal Processing*, 80:2067–2070.
- [29] P. Moulin and M. K. Mihcak. The data-hiding capacity of image sources. <http://www.ifp.uiuc.edu/~moulin/paper.html>, June 2001.
- [30] N. Nikolaidis and I. Pitas. Robust image watermarking in the spatial domain. *Signal Processing*, 66:385–403, 1998.
- [31] Proceedings. International workshop on information hiding.
- [32] M. Ramkumar and A.N. Akansu. Information theoretic bounds for data hiding in compressed images. *IEEE 2nd Workshop on Multimedia Signal Processing*, pages 267–272, Dec. 1998.
- [33] M. Ramkumar and A.N. Akansu. Theoretical capacity measures for data hiding in compressed images. *SPIE Multimedia Systems and Application*, 3528:482–492, 1998.
- [34] Regunathan Radhakrishnan, Nasir Memon. On the security of the SARI image authentication system. Proceedings of International Conference of Image Processing, Greece 2001.
- [35] S.D. Servetto, C.I. Podilchuk, and K. Ramachandran. Capacity issues in digital watermarking. *IEEE International Conf. on Image Processing*, 1:445–448, Oct. 1998.

- [36] C.E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 1948.
- [37] J.R. Smith and B.O. Comiskey. Modulation and information hiding in images. *Workshop on Information Hiding*, pages 463–470, 1996.
- [38] Special Issue. Watermarking. *IEEE Journal of Selected Areas in Communication*, May 1998.
- [39] Special Issue. Watermarking. *Proceedings of the IEEE*, 87(7), July 1999.
- [40] M. Swanson, B. Zhu, and A. Tewfik. Multiresolution video watermarking using perceptual models and scene segmentation. *Proc. IEEE Intl. Conf. on Image Processing*, II:558–561, 1997.
- [41] M. D. Swanson, M. Kobayashi, and A. H. Tewfik. Multimedia data-embedding and watermarking technologies. *Proceedings of the IEEE*, 86(6):1064–1087, June 1998.
- [42] A. Tewfik. White paper on data embedding, media annotation and copyright protection. *Cognicity Inc.*, Dec. 1997.
- [43] R.B. Wolfgang, C.I. Podilchuk, and E.J. Delp. Perceptual watermarks for digital images and video. *Proceedings of the IEEE*, 87(7):1108 –1126, July 1999.
- [44] P. Wong and N. Memon. Secret and Public Key Image Watermarking Schemes for Image Authentication and Ownership Verification. To appear in *IEEE Transactions on Image Processing*. October 2001.
- [45] W. Zhu, Z. Xiong, and Y. Q. Zhang. Multiresolution watermarking for images and video. *IEEE Transactions on Circuits and Systems for Video Technology*, 9(4):545–550, June 1999.

## 6 Figure Captions

Figure 1: An image with a visible watermark.

Figure 2: Watermark encoding process.

Figure 3: Watermark decoding process.

Figure 4: Meta-data tagging using information hiding.

Figure 5: Public key verification watermark insertion procedure.

Figure 6: Public key verification watermark extraction procedure.

Figure 7: SARI image authentication system - verification procedure

Figure 8: Random patterns and their smoothed versions used in Fridrich semi-fragile watermarking technique.

Figure 9: Tradeoffs in robust watermarking.

Figure 10: Example of a (a) binary iconic message and (b) message template.

Figure 11: Schematic of the watermark insertion process.

Figure 12: Schematic of the watermark extraction process.

Figure 13: a) Example of a watermarked image without rotation and scale transformation and its corresponding autocorrelation. (b) Image in top row after scale and rotation transformation and its corresponding autocorrelation.

Figure 14: Communication system model.

Figure 15: Watermarking as a communication system.

Figure 16: Discrete time additive channel noise model.

Figure 17: Multiplicative and additive watermarking channel noise model.