Robustness and Security of Digital Watermarks

Lesley R. Matheson, Stephen G. Mitchell*, Talal G. Shamoon, Robert E. Tarjan**, and Francis Zane***

STAR Lab
InterTrust Technologies Corporation
460 Oakmead Parkway
Sunnyvale, CA 94086.
{lrm,talal,mitchell,ret,fzane}@intertrust.com

Abstract. Digital watermarking is a nascent but promising technology that offers protection of unencrypted digital content. This paper is a brief technical survey of the multimedia watermarking landscape. The three main technical challenges faced by watermarking algorithms are fidelity, robustness and security. Current watermarking methods offer possibly acceptable fidelity and robustness against certain types of processing, such as data compression and noise addition, but are not sufficiently robust against geometric transforms such as scaling and cropping of images. Theoretical approaches have been developed that could lead to secure watermarking methods, but substantial gaps remain between theory and practice.

1 Introduction

The merging of computation and communication, as embodied for example in the Internet, offers substantial new opportunities for processing and distribution of valuable digital creations such as audio tracks, still images, and movies. At the same time, the new technology offers cheap and easy copying and distribution of pirated material. A standard and well-understood technical approach to reducing piracy is to use cryptography: valuable material is distributed in encrypted form, and only authorized users have the decryption keys. A complementary approach that offers protection of unencrypted material is *digital watermarking*.

This paper is a brief technical survey of the landscape of digital watermarking. Our goal is to understand the general principles that could lead to successful watermarking methods. Whereas cryptography is a relatively well-studied and stable field, serious study of digital watermarking began only recently, and much is not yet known. We begin in Section 2 by describing what we mean by watermarking, what content types might be marked, and what functionality watermarking might provide. In Section 3, we discuss what criteria should be used to

* and School of Electrical Engineering, Cornell University, Ithaca, NY 14853.
** and Department of Computer Science, Princeton University, Princeton, NJ 08544.
*** and Dept. of Computer Science, University of California at San Diego, La Jolla, CA 92093.

evaluate watermarking methods, and why successful watermarking might even be possible. In Section 4, we discuss the components of a generic watermarking system. In Section 5, we study the issue of robustness of watermarks to standard data processing, and in Section 6 we discuss various issues concerning watermark security. In Section 7, we study theoretical results about the resistance of watermarks to attacks. In Section 8 we offer a few concluding remarks. For a variety of interesting papers on watermarking and related topics, see [1, 33, 14].

2 What is Watermarking?

By *watermarking* we mean the embedding of encoded information into digital data so that the information is imperceptible, easily read by authorized parties only, and difficult to remove by unauthorized parties without destroying the (value of the) original data. We contrast this with several related, but distinct notions:

Steganography (hidden writing): Steganography is the imperceptible embedding of encoded information in data in a way that may or may not be robust, but with the assumption that a potential adversary is unaware of the existence of the hidden communication channel. Watermarking allows the possibility of an adversary knowing about the channel; ideally, we want methods resistant to malicious attack.

Visible watermarking: Here the mark is designed to be easily read by all parties, but this visibility may (or may not) spoil the original data. Examples of visible image watermarks include the glyph technology of Xerox [12] and a method of IBM [22].

Fragile watermarking: Here we embed information imperceptibly, but so that significant changes to the data destroy the watermark. A fragile watermark can serve as an embedded signature guaranteeing the authenticity of the data. Ideally, a fragile watermark might even reveal, through how it has been distorted, what processing the original data has undergone. Developing fragile watermarking methods is a promising research direction, but it is beyond our scope here.

In watermarking, it is important to distinguish between two broad content types. The first, *perceptual content*, includes audio tracks (speech and wideband), still images, and video clips. The second, *representational* or *abstract content*, includes natural language texts and programs written in general- or special-purpose programming languages. From the standpoint of watermarking, the main distinction between these two content types is the amount and kind of distortion each can tolerate. For perceptual content, it seems easier to make the distinction between "small" distortions of the data, as such those caused by a successful watermarking method, and "large," or value-destroying, distortions. Put differently, in the case of perceptual content, we are dealing with a continuous space of possibilities that has some relatively simple (though high-dimensional) geometry. What the geometry might be for various kinds of representational

content is largely unexplored terrain. Here we shall limit our attention to perceptual content, selecting our examples primarily from the realm of still images. Even for perceptual content, identifying the underlying geometry is a challenging problem.

The meaning of the watermark is another important issue. To understand the possible functions of watermarks, let us posit a simplified scenario in which a creator develops a piece of valuable content and sends it through a distribution channel, at the end of which it is released to various customers. The content can be marked in two places: by the creator, before it enters the distribution channel; and, by the distributor, as it leaves the channel. Marks introduced by the creator before distribution can depend only on the material itself, and not upon the recipient. Such watermarks can encode creator identity, copyright information, and content characteristics. These marks can be used to help defend copyright, to identify the creator for advertising and billing purposes, and to identify the content for use in metering and to facilitate database search.

Marking the material as it is released to a consumer affords the possibility of putting consumer identification or transaction information into the content, so that illegitimately-distributed content can be traced back to its point of release. The added flexibility allowed by transaction-based marking (sometimes called fingerprinting, although this word has also been used for other concepts) comes with some costs. First, the marking process must be efficient enough to not unduly impede the transaction. Preprocessing the content to make marking faster, or even pre-marking the content in many different ways and binding a specific mark to a specific consumer at transaction time, are ways to save time in transactions. Second, transactional marking raises the possibility of collusion attacks, in which pirates obtain several differently marked copies of the same material and combine all the copies to remove all the different marks. Most of the theoretical work on watermarking has been devoted to the problem of resisting collusive attacks. We look at this in Section 7.

For more complicated (and more realistic) distribution chains, there are more opportunities to introduce and to use watermarks. Assuming that the watermarking method supports multiple marking, the entire development and distribution history of a piece of content could be encoded in a sequence of successively embedded marks.

Some have suggested that watermarks should be human-readable [6]. We believe strongly that watermarks should be machine-readable, not (necessarily) human-readable. Machine-readable marks allow the possibility of active marking, in which marks are read and appropriate actions are taken in the course of content processing and distribution. A very simple example is copy control: a video player/copier outfitted with a watermark reader could seek a mark indicating "copyrighted material: no copying allowed," and disable the copying function when sensing such a mark. Another reason to use machine-readable marks is that they allow for much more efficient encoding of information into what is inherently a low-capacity communication channel.

3 Quality Criteria for Watermarking Methods

Critical criteria for watermarking methods include the following:

1. **Fidelity:** The changes entailed by marking should not affect the value of the content, and ideally the mark should be imperceptible. Specifically, experts in the medium should not be able to discriminate between the watermarked data and the original.
2. **Robustness:** Watermarks should survive standard data processing, such as would occur in a creation and distribution process. For still images, for example, such processing includes data compression, noisy transmission, digital-to-analog and analog-to-digital conversion (such as printing and scanning), color correction, sharpening and blurring, addition of captions, and geometric modifications such as cropping, scaling and rotation.
3. **Security:** Watermarks should survive deliberate attempts to remove them. Ideally, a watermark should remain readable up to the point where the content becomes modified enough to be of low value. A potential attacker can try standard processing techniques such as those mentioned above, but can also try less natural transformations specifically designed to erase watermarks. Attack becomes easier if the attacker has access to a watermark reader and can measure success. Attack also becomes easier if the attacker has access to differently marked versions of the same material.

Among these criteria, fidelity is usually paramount. The goal in building a successful marking method is to find a way to embed a mark with as much strength as possible (to provide robustness) while still preserving fidelity, by keeping the changes made by the mark under the perceptual threshold. We shall discuss this issue more fully in Section 5.

To obtain watermark security requires another key idea, that of *randomness*. The reason we might expect to be able to watermark perceptual data with some degree of security boils down to the following idea. Perceptual data has a very high number of places to put a mark (consider the number of pixels in a high-quality digital image.) The perceptual threshold allows small changes in a significant but relatively small number of such places. To spoil the mark successfully, an attacker who does not know the location of the mark must alter a large fraction of the places, thereby exceeding the perceptual threshold and destroying the original content. To guarantee that an attacker does not know the marked places, we can choose them randomly (or pseudo-randomly). This argument can be quantified, and (theoretically) gives security even in the presence of collusion. (See Section 7.)

Additional important criteria for watermarking methods include:

4. **Data capacity:** How many bits of information can the mark contain as a function of the size of the original content? How many marks can be added simultaneously?

5. **Accuracy of detection:** How accurately can the mark be read? What is the chance of a false positive (unmarked content appearing to have a mark) a false negative (marked data appearing to be unmarked), or a false reading (a mark misread as another mark)?
6. **Efficiency:** What are the computing time, storage requirements, and software or hardware size of the mark writing and reading processes? Are they real-time, so that they can be incorporated into playback or display mechanisms in an on-line setting? How do they interact with data compression and decompression?
7. **Data secrecy and storage requirements:** What information needs to be retained, or kept secret, about the marks, their meaning, and the marked material? Depending upon the watermarking method, such information can include encryption and decryption keys for computing and interpreting marks, a database mapping marks to their meanings, and a database containing (components of) original content that has been marked. A significant distinction here is between "original-based" watermarking methods, in which the original data is required to read the marks, and "no-original" methods, in which marks can be read without having the original. No-original algorithms are much more flexible and useful but harder to make robust. Many of the early algorithms in the literature are original-based methods, which have restricted practical functionality.

4 The Components of a Watermarking System

There are three components to a watermarking system: a watermark writing algorithm, a watermark reading algorithm, and a database (or databases) to store needed information about marks written and data that has been marked. We shall discuss watermark writing and reading from a generic point of view. When specifics or examples are needed to make concepts concrete, we shall select them from the domain of still images, although the same principles apply to watermarking audio tracks, video clips or similar kinds of data. Since database technology is relatively well-understood, we shall not comment on this component of a watermarking system, except to mention information that might have to be stored in such a database.

The first step in watermark writing is to choose a representation of the original data. An important property of media data is that it is high-dimensional: think of the number of pixels in a high-quality image. The representation may be the original representation (pixels for images); or it may be a transformed representation, such as Fourier components, discrete cosine components, or wavelet components; or some higher-level representation, such as an object or feature representation. If a transformed representation is used, it may be applied to the entire data (the whole image), or on a block-by-block basis. For images, watermarks have been applied to pixels [2], Fourier components [24], whole-image and block-based discrete cosine components [8, 23, 39], wavelet components [23], and Fourier-Mellin components [25, 13]. One use of a transformed representation

is to make the components of the data more independent; pixels in an image, for example, are highly correlated locally, which is not true of discrete cosine components.

Once a representation has been determined, a subset of the components must be chosen to be marked. This choice is generally made with the goal of preserving fidelity while enhancing robustness and security as much as possible. One idea in the literature is to mark the largest-magnitude discrete cosine components [8], the justification being that these components are perceptually significant and likely to be preserved by common data processing techniques, such as data compression. Though this may be a good approach for an original-based watermarking scheme, it may not be so useful for a no-original scheme, for reasons discussed below. For typical data, the largest magnitude components are mostly the low-frequency ones, and marking low frequencies may serve just as well. A possibly better idea for a no-original scheme is to mark mid-frequency components [13, 37, 19] with the hope that these components are perceptually significant but relatively low-energy. Herigal, et al. [13] for example, mark in the Fourier domain. They avoid "the largest (high energy) components (at about the lowest 10% of the frequencies)" and use "components at the medium frequencies (about next 30%)." They also suggest the possibility of "marking the largest components (inside the allowed frequency range)."

Another guide to choosing components is to seek some that are invariant to certain kinds of processing. For example, one may choose to represent a color image using a luminosity-chrominance basis, and mark only the luminosity components, thereby rendering the watermark robust to a color-to-grayscale transformation. A similar idea discussed more fully in the next section is to mark certain Fourier-Mellin components of an image [25, 13], thereby obtaining some robustness to the geometric transformations of scaling, cropping and rotation. To enhance security, one may choose to mark only a random subset of the set of components selected for robustness.

Having chosen components (places) to mark, one must choose the mark values and combine them with the chosen components to obtain modified components, which replace the original components and are used to construct a modified copy of the original data. The mark values may be an encoded and possibly encrypted representation of the information to be conveyed by the mark, or they may be chosen randomly and merely associated with the intended information (via a database entry). It is worth noting that encryption by itself will serve the purpose of making the watermark values appear random. A common choice of mark values is $\{0, 1\}$ or $\{-1, +1\}$, although we shall see later that security needs dictate other choices. To deal with the issue of perceptibility, we may choose to multiply each watermark value by a strength parameter, which may be globally chosen or may depend on the particular component being marked and on the particular data being marked: sophisticated marking algorithms use perceptual masking models to choose strength parameters [39, 40, 31], seeking maximum-strength marks that lie within the perceptual threshold.

As a way of combining mark values with original component values, we distinguish between *addition*, in which each strengthened mark value is added to the corresponding original component value, and *replacement*, in which each strengthened mark value replaces the corresponding original component value. Other ways of combining values can generally be reduced to addition by an appropriate transformation. For example, a multiplicative marking scheme can be reduced to an additive marking scheme by applying a logarithmic transformation.

We can represent a generic additive watermark-writing method symbolically as follows. Assume that n components of the data are to be marked, and that the original component values are $d_1, d_2 \ldots, d_n$. Let $w_1, w_2 \ldots, w_n$ be the selected watermark values, and let $s_1, s_2 \ldots, s_n$ be the desired watermark strengths. Then the watermark writing process consists of replacing each $d_i, 1 \leq i \leq n$, by $d'_i = d_i + s_i w_i$. The corresponding replacement watermark-writing method would instead replace each d_i by $d'_i = s_i w_i$. If there is a finite range to each component (as for example with pixels), we must truncate each d'_i to keep it in range.

The second major component of a watermarking system is the watermark reader, which of course must match the writer. To read a mark, we first transform the data into the representation used for mark writing. Then we extract the components $d^*_1, d^*_2, \ldots, d^*_n$ that correspond to the ones that were marked. In a replacement-based scheme, these values should be approximately the strengthened mark values $s_1 w_1, s_2 w_2 \ldots, s_n w_n$. We can merely divide each d^*_i by the corresponding strength s_i and attempt to interpret the resulting vector d^*_1/s_1, $d^*_2/s_2, \ldots, d^*_n/s_n$ as a mark. One way to make the reading process robust is to use an error-correcting code in choosing marks and interpreting them. Another way is to apply signal detection theory [32] (see also [37, 20, 19]) and do a correlation-based hypothesis test. Namely, we compute a correlation (a dot product normalized in some way) between the hypothetical watermark $d^*_1/s_1, d^*_2/s_2, \ldots, d^*_n/s_n$ and an actual watermark $w_1, w_2, \ldots, w_n$, and conclude that the latter is present in the data if the correlation exceeds some threshold.

Virtually the same methods can be used to read additive watermarks. The connection is tightest for original-based additive marking. If we have access to the appropriate components, $d_1, d_2, \ldots, d_n$ of the original data, we can compute a hypothetical watermark by first subtracting these components and then dividing by the strengths: $w^*_i = (d^*_i - d_i)/s_i$. We can then apply error correction or a correlation test to the sequence $w^*_1, w^*_2, \ldots, w^*_n$ to attempt to match it against an actual mark. The CKLS original-based watermarking method [8] uses such a correlation test to read the mark.

Reading a no-original additive watermark is more problematic. Fortunately, the correlation-based method still works if we merely correlate the reduced-strength components $d^*_1/s_1, d^*_2/s_2, \ldots d^*_n/s_n$ with an actual watermark w_1, w_2, $\ldots$, w_n and apply a threshold test [37, 20, 19]. A corresponding no-original version of the CKLS method is described in [30]. Such an approach works because the reduced-strength components d^*_i/s_i are approximately the watermark values w_i

plus the reduced-strength original components d_i/s_i, and the correlation between the w_i and the d_i/s_i is approximately zero, but with high variance.

A major hurdle in no-original watermarking is to reduce the noise in the detection process caused by the presence of the original data when doing watermark reading [37, 19]. One way to accomplish this is to mark low-energy but still significant components, such as middle-frequency components as mentioned above. Subtracting out the original, when this is possible, can be viewed as just a very powerful noise-reduction technique applied to standard correlation-based signal detection.

5 Robustness

As discussed briefly in Section 3, to be robust, a watermark must survive two types of standard processing techniques: *alignment-preserving transformations*, which include data compression, quantization, data conversion (digital-to-analog and analog-to-digital conversion), and others; and *alignment-altering transformations*, such as (in the case of images), cropping, scaling and rotation (Data conversion with severe distortion or imprecise resampling may actually be alignment-altering rather than alignment-preserving.) The current state-of-the-art is that there are a variety of similar watermarking algorithms for various media types that survive alignment-preserving tranformations reasonably well. Many of these algorithms use frequency-based representations and rely on some kind of perceptual model to embed a maximum-strength imperceptible mark (e.g. [40]).

Lacy, et al. [18] argue that *compressed* data, not the original *baseband* (raw or uncompressed) data, is what should be protected. They propose an audio watermarking algorithm that is tightly integrated with a perceptual audio data compressor. Such an algorithm allows reading a watermark from the compressed data, a capability that may be a requirement in on-line transaction-based systems. Whether such watermarks survive decompression remains to be tested empirically.

Surviving alignment-altering transformations is problematic. For example, devising still-image watermarking methods that are robust to scaling, cropping and rotation is a challenging problem, especially for combinations of these transformations as would occur, for example, in creating a photomontage. Several approaches exist:

1. In an original-based method, one can align a transformed watermarked image against the original, using standard registration or pattern-matching methods.
2. In a no-original based method, one can add a universal *registration* mark and align a transformed marked image against the registration mark.
3. In a no-original method, one can attempt to do a self-alignment of a transformed image, based on some set of distinguishable features.
4. As mentioned in Section 3, one can put the watermark into a set of components that are invariant to certain transformations. For still images the

magnitudes of Fourier-Mellin components are invariant under translation, rotation, and scaling (in an abstract, continuous setting) [5, 36].

Ó Ruanaidh and Pun [25] have explored Method 4, that of marking the magnitudes of the Fourier-Mellin components, as a way to make watermarks robust against translation, rotation and scaling. This idea has been refined by Herigal, et al. [13]. They first take the logarithm of the luminance levels (to match the human visual system) and then do a Fourier transform. They mark the magnitudes of medium-frequency components. Additionally, they add a registration template based on a Fourier-Mellin transform of the magnitudes of the Fourier components. This template is intended to be robust against rotation and scaling. Their approach combines Methods 2 and 4. It is worth noting that taking the logarithm as the first step has the effect of increasing the watermark signal-to-noise ratio and hence of making detection easier.

6 Security

Security of watermarks is receiving increasing attention, especially from the academic community. A variety of attacks on various kinds of watermarking schemes have been proposed and studied [26, 20, 38, 15, 21, 10]. There are a variety of issues concerning watermark security that are properly in the domain of cryptography and cryptographic protocols, and these should be considered separately from the issue of whether marks can be erased. For example, by using standard cryptographic methods, one can guarrantee (up to the security of the cryptographic scheme) that watermarks cannot be read or forged by unauthorized parties, although preventing false claims of ownership may require an appropriate information registry. Craver, et al. [10] observed that additive original-based schemes such as the CKLS algorithm [8] can be subjected to a forgery attack in which a forger creates his or her own watermark and subtracts it from a previously marked original, creating a fake "original" that the forger claims to own. Again, cryptographic techniques are the appropriate way to guard against attacks like this.

Turning to attacks designed to make watermarks unreadable, it is well-known that a simple least-significant-bit scheme can be defeated by randomizing the least-significant bits that contain the watermark, or by setting all these bits to zero. Schemes based on perceptual modeling that attempt to insert maximum-power watermarks are much harder to attack, but small amounts of scaling and cropping will erase many kinds of watermarks. For example, Kilian, et al. [17] observed that the CKLS mark can be rendered unreadable by cropping a few rows and columns of pixels and scaling the image to the original size. Such an attack can be countered by aligning (or registering) to the original image; or, if the original is not available, by aligning to a previously inserted registration mark, or by using a watermark that is robust to such transformations, as discussed in Section 5.

Making watermark writers and readers publicly available creates security risks, even if the algorithms are black boxes. For example, a watermark scheme

that uses a universal, additive registration mark can be made unreadable by taking the negative of the marked data and remarking it, thereby subtracting out the registration mark (Bill Horne, oral communication, 1997). A scheme that uses a public reader may be susceptible to a sensitivity-analysis attack such as described by Cox and Linnartz [9] and analyzed by Linnartz and Van Dijk [20]. Fridrich, et al. [11] have proposed a watermarking method that uses key-dependent basis functions and may allow the construction of a secure public reader.

A particularly potent kind of attack is a *collusion attack*, in which an attacker obtains several differently marked copies of the same data, or several different pieces of data marked in the same way. Kilian, et al. [17] have observed that correlative-reader watermarks that use a componentwise $\{-1, 1\}$ or $\{0, 1\}$ distribution (common in the literature) or even a component-wise uniform distribution are at risk of attack with as few as five or six differently-marked copies. Resistance to collusive attack is the main focus of the theoretical work we discuss in the next section.

7 Models of Security

A body of work exists devoted to answering the question of whether truly secure watermarks can exist, and what the characteristics of such marks might be. Most of this work deals with resistance to collusive attacks. Such work is necessarily theoretical and relies on modeling assumptions. A key issue is the extent to which the emerging theories can be applied to practice.

An early and intriguing paper is that of Wagner [43], who proposed the use of randomly selected additive watermarks and did a preliminary study of the resistance of such marks to collusive attacks. Blakely, et al. [3] looked at a combinatorial model for collusion resistance, and proposed a scheme that offers k-way collusion resistance within the model but requires a number of watermark bits exponential in the number of colluders. Chor, et al. [7] worked on a related problem involving tracing pirates in a broadcast distribution system using a multiple-key protocol. Boneh and Shaw [4] combined the Chor, et al. work with a simple collusion-resistant scheme to yield a watermarking method that provides defense against k-way collusion and conveys b bits of information in a watermark of size $O(k^4 b)$ bits. Follow-on work to that of Chor, et al. and Boneh and Shaw appears in [28, 29, 27].

The Boneh-Shaw model is a discrete combinatorial framework that captures the notion of collusion resistance, but ignores other issues of watermark security and robustness. They posit a watermark consisting of n positions, each position selected from an alphabet of size s. An attacker in posession of k differently watermarked copies is allowed to spoil any position in which two of the obtained copies differ. The goal of the watermarker is to identify at least one of the watermarks using only the information contained in the positions in which all k watermarks are the same.

Boneh and Shaw do not address the question of how to provide individual robust marking positions (which cannot be attacked unless detected by difference analysis) nor do they consider the possibility that marked positions, even if detected, might be difficult to spoil (for perceptual or other reasons) This makes their method more suited for representational-content watermarking (see Section 2) than for perceptual-content watermarking.

A model of the latter kind of watermarking has been investigated by Kilian, et al. [17], building on ideas of J. Kilian and F.T. Leighton (oral communications, 1996) and a draft set of notes of Leighton (1996). The model assumes that the original data is an n-dimensional vector, with each component independently distributed according to $N(0,1)$, the Gaussian distribution with mean zero and variance 1. The model further assumes a perceptual threshold based on Euclidean distance. For an additive watermark that is also an n-dimensional vector with independent components distributed according to $\epsilon N(0,1)$ (for a suitable choice of ϵ depending on the perceptual threshold) and a correlative original-based watermark reader, Kilian, et al. proved that the watermarks can carry $O(n/k^2)$ bits of information while resisting k-way collusive attacks with high probability. Equivalently, $O(k^2 b)$ watermark components are needed to carry b bits of information while being secure against k-way collusive attack. Further work by Ergun, Kilian and Kumar (unpublished notes, 1997), refined and tightened by Mitchell, Tarjan and Zane (unpublished notes, 1997) has shown that, within the Kilian-Leighton statistical model, no watermarking method can offer better resistance to collusive attacks. Specifically, collusive attack based on averaging and addition of Gaussian noise will erase any watermark with high probability, given $\Omega(n/k^2)$ differently marked copies, if n is the dimension of the watermarks.

Related but independent work has been done by Karakos and Papamarcou [16]. They consider the ability of maximum-strength watermarks with original-based correlative reading to withstand the attack of averaging plus addition of Gaussian noise, within a Euclidean perceptual threshold model. They consider only a single-copy attack and a two-copy attack. They show that such watermarks can convey up to 0.5 bits of information per dimension while being secure against a one-copy attack, and up to 0.146 bits per dimension while being secure against a two-copy attack.

One reason why no-original watermarking works, at least theoretically, is that in high dimensions randomly selected watermarks are, with high probability, almost orthogonal to the data and to each other [20, 37]. Tirkel et al. [42, 34, 35] discuss the issue of orthogonality at length, and propose the construction of watermarks that are exactly orthogonal to each other, but this may be an unnecessary step in practice. (They use pseudo-random bit sequences as watermarks rather than Gaussian noise.) Swanson, et al. [41] suggest a scheme that chooses a random watermark direction and encodes a mark in the hidden direction, with a strength determined by a perceptual model.

Some extensions of the Kilian, et al. results are possible. First, it is straightforward to extend the result to the no-original setting; only the constant factors

change. Second, the watermark need not be Gaussian, but can be maximum-strength (or fixed strength) with randomly chosen direction within the Euclidean threshold model, because, for high dimensions, Gaussian and random-direction fixed-strength watermarks behave approximately the same. Thus the Swanson et al. marking algorithm falls within this theory.

Much remains to be done to extend this theoretical work and to determine if it has any practical relevance. The Kilian, et al. Euclidean perceptual model breaks down in reality because small geometric distortions can produce large changes in Euclidean distance. Also, a correlative reader must compute a correlation for each possible watermark, leading to a computation that is exponential in the number of bits of information conveyed. One way to improve the efficiency of the reader is to use a small set of signalling patterns, either combined in all the components or distributed over subsets of components. A third direction to study is the relationship between the combinatorial and statistical models, and to determine whether they can usefully be combined. Finally, the Boneh-Shaw bound on watermark size is $O(k^4 b)$ to protect against k colluders and convey b bits. Reducing this bound, or proving it tight, is an open problem.

8 Remarks

Digital watermarking, though young, is a rapidly expanding field. It combines elements of cryptography, signal processing, information theory, coding theory, probability and statistics, game theory and other disciplines. Whether all the activity in this area will lead to robust, practical watermarking schemes remains to be seen, but certainly the field is full of exciting possibilities.

References

1. R. Anderson, Ed., *Information Hiding, First International Workshop Proceedings*, *Lecture Notes in Computer Science* 1174, Springer-Verlag, Berlin, 1996.
2. W. Bender, D. Gruhl, N. Marimoto, and A. La, "Techniques for data hiding," *IBM Systems Journal* 35 (1996).
3. G.R. Blakely, C. Meadors, and G.B. Purdy, "Fingerprinting long unforgiving messages," *Crypto '85, Lecture Notes in Computer Science* 218, Springer-Verlag, Berlin (1985), pp.180-189.
4. D. Boneh and J. Shaw, "Collusion-secure fingerprinting for digital data," *Crypto '95, Lecture Notes in Computer Science* 963, Springer-Verlag, Berlin, 1995, pp. 452–465.
5. R.D. Brandt and F. Lin, "Representations that uniquely characterize images modulo translation, rotation, and scaling," *Pattern Recognition Letters* 17 (1996), pp. 1001-1015.
6. G. W. Braudaway, "Protecting publicly-available images with an invisible image watermark," *Proc. IEEE Int. Conf. on Image Procesing, ICIP-97* (1997), Vol. I, pp. 524-51.
7. B. Chor, A. Fiat, and M. Naor, "Tracing traitors," *Crypto '94, Lecture Notes in Computer Science* 963, Springer-Verlag, Berlin, 1995, pp. 452-465.

8. I. Cox, J. Kilian, T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. on Image Processing* 6 (1997), 1673-1687.
9. I. Cox and J.-P. Linnartz, "Public watermarks and resistance to tampering," *Proc. IEEE Conf. on Image Processing* (1997), CD-ROM.
10. S. Craver, N. Memon, B.L. Yeo, and M. Yeung, "Can invisible watermarks resolve rightful ownerships?", *IBM Research Report* RC 20509 (1996).
11. J. Fridrich, A.C. Baldoza, and R.J. Simard, "Robust digital watermarking based on key-dependent basis functions," *Preliminary Proc. Second International Information Hiding Workshop* (1998).
12. D.L. Hecht, " Embedded data glyph technology for hardcopy digital documents," *SPIE* 2171 (1995).
13. A. Herrigal, J.J Ó Ruanaidh, W. Peterson, S. Pereira, and T. Pun, "Secure copyright protection techniques for digital images," *Preliminary Proceedings of the Second International Information Hiding Workshop* (1998).
14. *Proceedings of the IEEE International Conference on Image Processing, ICIP-47,* Vols. I-III, IEEE Computer Society, Los Alamitos, CA, 1997.
15. N.F. Johnson and S. Sajodia, "Steganalysis of images created using current steganography software," *Preliminary Proc. Second International Information Hiding Workshop* (1998).
16. D. Karakos and A. Papamarcou, "Some results on the information capacity of authentication channels," Dept. of Electrical Engineering, University of Maryland, College Park, MD, manuscript, 1997.
17. J. Kilian, F.T. Leighton, L. Matheson, T. Shamoon, R. Tarjan, and F. Zane, "Resistance of digital fingerprints to collusional attacks," unpublished manuscript (1997).
18. J. Lacy, S.R. Quackenbush, A. Reibman, and J.H. Snyder, "Intellectual property protection systems and digital watermarking," *Preliminary Proceedings of the Second International Information Hiding Workshop* (1998).
19. J. P. Linnartz, T. Kalker, and G. Depovere, "Modelling the false alarm and missed detection rate for electronic watermarks," *Preliminary Proceedings of the Second International Information Hiding Workshop* (1998)
20. J. P. Linnartz and M. Van Dijk, "Analysis of the sensitivity attack against electronic watermarks in images," *Preliminary Proceedings of the Second International Information Hiding Workshop* (1998).
21. M. Maes, "Twin peaks: the histogram attack on fixed depth image watermarks," *Preliminary Proc. Second International Information Hiding Workshop* (1998).
22. K.A.Magerlein, G.W. Braudaway, and F.C. Mintzer, "Protecting publically-available images with a visible image watermark," *Proc. SPIE Conf. on Optical Security and Counterfeit Deterrence Techniques*, SPIE 2659 (1996), pp. 126-132.
23. J.J.K Ó Ruanaidh, W. J. Dowling, and F. M. Boland, "Watermarking digital images for copyright protection," *IEE Proc. on Vision, Image and Signal Processing* 143 (1996), pp. 250-256.
24. J.J.K. Ó Ruanaidh, W.J. Dowling, and F.M. Boland, "Phase watermarking of digital images," *Proc. IEEE International Conference on Image Processing ICIP-96* (1996), pp. 239-242.
25. J. Ó Ruanaidh and T. Pun, "Rotation, translation and scale invariant digital image watermarking," *Proceedings 1997 IEEE International Conference on Image Processing* (1997), Vol. I, pp. 536-539.
26. F. Peticolas, R. Anderson, and M. Kuhn, "Attacks on copyright marking systems," *Preliminary Proc. Second International Information Hiding Workshop* (1998).

27. B. Pfitzmann and M. Shaunter, "Anonymous fingerprinting, (extended abstract)," *EUROCRYPT '96, Lecture Notes in Computer Science* 1070, Springer-Verlag, Berlin (1996), pp. 84-95.
28. B. Pfitzmann and M. Waidner, "Asymmetric fingerprinting for larger collusions," *4th ACM Conf. on Computer and Communications Security* (1997), pp. 151-160.
29. B. Pfitzmann and M. Waidner, "Anonymous fingerprinting," IBM Research Report RZ 2221 (1996).
30. A. Piva, M. Barni, F. Bartolini, and V. Cappellini, "DCT-based watermark recovering without resorting to the uncorrupted original image," *Proceedings 1997 IEEE International Conference on Image Processing* (1997) Vol. I, pp. 520-523.
31. C. Podilchuck and W. Zeng, "Digital image watermarking using visual models," *IS&T/SPIE Electronic Imaging* 3016 (1997).
32. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd Edition, Springer-Verlag, New York, 1994.
33. *Preliminary Proceedings of the Second International Information Hiding Workshop,* 1998.
34. R. van Schyndel, A. Tirkel, and C. Osborne, "A digital watermark," *Proceedings 1994 IEEE International Conference on Image Processing*, (1994), pp. 86-90.
35. R. van Schyndel, A. Tirkel, and C. Osborne, "Towards a robust digital watermark," *Proceedings DICTA-95* (1993), pp. 378-385.
36. Y. Sheng and H.H. Arsenault, "Experiments on pattern recognition using invariant Fourier-Mellon descriptors," *J. Optical Society of America A* 3 (1986), pp. 771-776.
37. J.R. Smith and B.O. Comisky, "Modulation and information hiding in images," *Information Hiding, First International Workshop Procedings*, R. Anderson, ed., *Lecture Notes in Computer Science* 1174, Springer-Verlag, Berlin (1996), pp. 207-226.
38. S. Sowers and A. Youssef, "Testing digital watermark resistance to destruction," *Preliminary Proc. Second International Information Hiding Workshop* (1998).
39. M.D.Swanson, B. Zhu, and A.A. Tewfik, "Transparent robust image watermarking," *Proc IEEE Int. Conf. On Image Processing, ICIP-96* Vol. 3, (1996) pp. 211-214.
40. M.D. Swanson, B. Zhu, and A.H. Tewfik, "Robust image watermaking using perceptual models," unpublished manuscipt (1997).
41. M. Swanson, B. Zhu, and A. Tewfik, "Data Hiding for Video in Video," *Proceedings 1997 IEEE International Conference on Image Processing*, (1997), Vol II, pp. 676-679.
42. A. Tirkel, G. Rankin, R. van Schyndel, W. Ho, N. Mee, and C Osborne, "Electronic watermark," *Proceedings DICTA-93* (1993), pp. 666-673.
43. N. Wagner, "Fingerprinting," *Proc. 1983 IEEE Symp. on Security and Privacy* (1983), pp. 18-22.
44. W. Zeng and B. Liu, "On resolving rightful ownerships of digital images by invisible watermarks," *Proc. IEEE Int. Conf. on Image Processing, ICIP-97* (1997), Vol. I, pp. 552-555.