Secure Spread Spectrum Watermarking for Multimedia

Ingemar J. Cox, *Senior Member, IEEE,* Joe Kilian, F. Thomson Leighton, and Talal Shamoon, *Member, IEEE*

***Abstract*—This paper presents a secure (tamper-resistant) algorithm for watermarking images, and a methodology for digital watermarking that may be generalized to audio, video, and multimedia data. We advocate that a watermark should be constructed as an independent and identically distributed (i.i.d.) Gaussian random vector that is imperceptibly inserted in a spread-spectrum-like fashion into the perceptually *most* significant spectral components of the data. We argue that insertion of a watermark under this regime makes the watermark robust to signal processing operations (such as lossy compression, filtering, digital-analog and analog-digital conversion, requantization, etc.), and common geometric transformations (such as cropping, scaling, translation, and rotation) provided that the original image is available and that it can be succesfully registered against the transformed watermarked image. In these cases, the watermark detector unambiguously identifies the owner. Further, the use of Gaussian noise, ensures strong resilience to multiple-document, or collusional, attacks. Experimental results are provided to support these claims, along with an exposition of pending open problems.**

***Index Terms*— Intellectual property, fingerprinting, multimedia, security, steganography, watermarking.**

I. Introduction

THE PROLIFERATION of digitized media (audio, image, and video) is creating a pressing need for copyright enforcement schemes that protect copyright ownership. Conventional cryptographic systems permit only valid keyholders access to encrypted data, but once such data is decrypted there is no way to track its reproduction or retransmission. Therefore, conventional cryptography provides little protection against data piracy, in which a publisher is confronted with unauthorized reproduction of information. A digital watermark is intended to complement cryptographic processes. It is a visible, or preferably invisible, identification code that is permanently embedded in the data and remains present within the data after any decryption process. In the context of this work, data refers to audio (speech and music), images (photographs and graphics), and video (movies). It does not include ASCII representations of text, but does include text represented as an image. Many of the properties of the scheme presented in this work may be adapted to accommodate audio and video implementations, but the algorithms here specifically apply to images.

A simple example of a digital watermark would be a visible "seal" placed over an image to identify the copyright owner (e.g., [2]). A visible watermark is limited in many ways. It marrs the image fidelity and is susceptible to attack through direct image processing. A watermark may contain additional information, including the identity of the purchaser of a particular copy of the material. In order to be effective, a watermark should have the characteristics outlined below.

Unobtrusiveness: The watermark should be perceptually invisible, or its presence should not interfere with the work being protected.

Robustness: The watermark must be difficult (hopefully impossible) to remove. If only partial knowledge is available (for example, the exact location of the watermark in an image is unknown), then attempts to remove or destroy a watermark should result in severe degradation in fidelity before the watermark is lost. In particular, the watermark should be robust in the following areas.

- *Common signal processing:* The watermark should still be retrievable even if common signal processing operations are applied to the data. These include, digital-to-analog and analog-to-digital conversion, resampling, requantization (including dithering and recompression), and common signal enhancements to image contrast and color, or audio bass and treble, for example.
- *Common geometric distortions (image and video data):* Watermarks in image and video data should also be immune from geometric image operations such as rotation, translation, cropping and scaling.
- *Subterfuge attacks (collusion and forgery):* In addition, the watermark should be robust to collusion by multiple individuals who each possess a watermarked copy of the data. That is, the watermark should be robust to combining copies of the same data set to destroy the watermarks. Further, if a digital watermark is to be used in litigation, it must be impossible for colluders to combine their images to generate a different valid watermark with the intention of framing a third party.

Manuscript received January 14, 1996; revised January 24, 1997. Portions of this work were reprinted, with permission, from the Proceedings of the IEEE Conference on Image Processing, 1996, and from the Proceedings of the First International Conference on Data Hiding (Springer-Verlag, 1996). The associate editor coordinating the reivew of this manuscript and approving it for publication was Prof. Sarah Rajala.

I. J. Cox and J. Kilian are with NEC Research Institute, Princeton, NJ 08540 USA (e-mail: ingemar@research.nj.nec.com; joe@research.nj.nec.com).

F. T. Leighton is with the Mathematics Department and Laboratory for Computer Science, The Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: ftl@math.mit.edu).

T. Shamoon is with InterTrust STAR Laboratory, Sunnyvale, CA 94086 USA (e-mail: talal@intertrust.com).

Publisher Item Identifier S 1057-7149(97)08460-1.

Universality: The same digital watermarking algorithm should apply to all three media under consideration. This is potentially helpful in the watermarking of multimedia products. Also, this feature is conducive to implementation of audio and image/video watermarking algorithms on common hardware.

Unambiguousness: Retrieval of the watermark should unambiguously identify the owner. Furthermore, the accuracy of owner identification should degrade gracefully in the face of attack.

There are two parts to building a strong watermark: the *watermark structure* and the *insertion strategy*. In order for a watermark to be robust and secure, these two components must be designed correctly. We provide two key insights that make our watermark both robust and secure: We argue that the watermark be placed explicitly in the perceptually most significant components of the data, and that the watermark be composed of random numbers drawn from a Gaussian $(N(0,1))$ distribution.

The stipulation that the watermark be placed in the perceptually significant components means that an attacker must target the fundamental structural components of the data, thereby heightening the chances of fidelity degradation. While this strategy may seem counterintuitive from the point of view of steganography (how can these components hide any signal?), we discovered that the significant components have a *perceptual capacity* that allows watermark insertion without perceptual degradation. Further, most processing techniques applied to media data tend to leave the perceptually significant components intact. While one may choose from a variety of such components, in this paper, we focus on the perceptually significant *spectral* components of the data. This simultaneously yields high perceptual capacity and achieves a uniform spread of watermark energy in the pixel domain.

The principle underlying our watermark structuring strategy is that the mark be constructed from independent, identically distributed (i.i.d.) samples drawn from a Gaussian distribution. Once the significant components are located, Gaussian noise is injected therein. The choice of this distribution gives resilient performance against collusion attacks. The Gaussian watermark also gives our scheme strong performance in the face of quantization, and may be structured to provide low false positive and false negative detection. This is discussed below, and elaborated on in [13].

Finally, note that the techniques presented herein do not provide proof of content ownership on their own. The focus of this paper are algorithms that insert messages into content in an extremely secure and robust fashion. Nothing prevents someone from inserting another message and claiming ownership. However, it is possible to couple our methods with strong authentication and other cryptographic techniques in order to provide complete, secure and robust owner identification and authentication.

Section III begins with a discussion of how common signal transformations, such as compression, quantization, and manipulation, affect the frequency spectrum of a signal. This discussion motivates our belief that a watermark should be embedded in the data's perceptually significant frequency components. Of course, the major problem then becomes how to imperceptibly insert a watermark into perceptually significant components of the frequency spectrum. Section III-A proposes a solution based on ideas from spread spectrum communications. In particular, we present a watermarking algorithm that relies on the use of the original image to extract the watermark. Section IV provides an analysis based on possible collusion attacks that indicates that a binary watermark is not as robust as a continuous one. Furthermore, we show that a watermark structure based on sampling drawn from multiple i.i.d Gaussian random variables offers good protection against collusion. Ultimately, no watermarking system can be made perfect. For example, a watermark placed in a textual image may be eliminated by using optical character recognition technology. However, for common signal and geometric distortions, the experimental results of Section V suggest that our system satisfies most of the properties discussed in the introduction, and displays strong immunity to a variety of attacks in a collusion resistant manner. Finally, Section VI discusses possible weaknesses and potential enhancements to the system and describes open problems and subsequent work.

II. Previous Work

Several previous digital watermarking methods have been proposed. Turner [25] proposed a method for inserting an identification string into a digital audio signal by substituting the "insignificant" bits of randomly selected audio samples with the bits of an identification code. Bits are deemed "insignificant" if their alteration is inaudible. Such a system is also appropriate for two-dimensional (2-D) data such as images, as discussed in [26]. Unfortunately, Turner's method may easily be circumvented. For example, if it is known that the algorithm only affects the least significant two bits of a word, then it is possible to randomly flip *all* such bits, thereby destroying any existing identification code.

Caronni [6] suggests adding *tags*—small geometric patterns—to digitized images at brightness levels that are imperceptible. While the idea of hiding a spatial watermark in an image is fundamentally sound, this scheme may be susceptible to attack by filtering and redigitization. The fainter such watermarks are, the more susceptible they are such attacks and geometric shapes provide only a limited alphabet with which to encode information. Moreover, the scheme is not applicable to audio data and may not be robust to common geometric distortions, especially cropping.

Brassil *et al.* [4] propose three methods appropriate for document images in which text is common. Digital watermarks are coded by 1) vertically shifting text lines, 2) horizontally shifting words, or 3) altering text features such as the vertical endlines of individual characters. Unfortunately, all three proposals are easily defeated, as discussed by the authors. Moreover, these techniques are restricted exclusively to images containing text.

Tanaka *et al.* [19], [24] describe several watermarking schemes that rely on embedding watermarks that resemble quantization noise. Their ideas hinge on the notion that quantization noise is typically imperceptible to viewers. Their first scheme injects a watermark into an image by using a predetermined data stream to guide level selection in a predictive quantizer. The data stream is chosen so that the resulting image looks like quantization noise. A variation on this scheme is also presented, where a watermark in the form of a dithering matrix is used to dither an image in a certain way. There are several drawbacks to these schemes. The most important is that they are susceptible to signal processing, especially requantization, and geometric attacks such as cropping. Furthermore, they degrade an image in the same way that predictive coding and dithering can.

In [24], the authors also propose a scheme for watermarking facsimile data. This scheme shortens or lengthens certain runs of data in the run length code used to generate the coded fax image. This proposal is susceptible to digital-to-analog and analog-to-digital attacks. In particular, randomizing the least significant bit (LSB) of each pixel's intensity will completely alter the resulting run length encoding. Tanaka *et al.* also propose a watermarking method for "color-scaled picture and video sequences". This method applies the same signal transform as the Joint Photographers Expert Group (JPEG) (discrete cosine transform of 8 × 8 subblocks of an image) and embeds a watermark in the coefficient quantization module. While being compatible with existing transform coders, this scheme may be susceptible to requantization and filtering and is equivalent to coding the watermark in the LSB's of the transform coefficients.

In a recent paper, Macq and Quisquater [18] briefly discuss the issue of watermarking digital images as part of a general survey on cryptography and digital television. The authors provide a description of a procedure to insert a watermark into the least significant bits of pixels located in the vicinity of image contours. Since it relies on modifications of the least significant bits, the watermark is easily destroyed. Further, their method is restricted to images, in that it seeks to insert the watermark into image regions that lie on the edge of contours. Bender *et al.* [3] describe two watermarking schemes. The first is a statistical method called *patchwork*. Patchwork randomly chooses n pairs of image points, (a_i, b_i), and increases the brightness at a_i by one unit while correspondingly decreasing the brightness of b_i. The expected value of the sum of the differences of the n pairs of points is then $2n$, provided certain statistical properties of the image are true.

The second method is called "texture block coding," wherein a region of random texture pattern found in the image is copied to an area of the image with similar texture. Autocorrelation is then used to recover each texture region. The most significant problem with this technique is that it is only appropriate for images that possess large areas of random texture. The technique could not be used on images of text, for example, nor is there a direct analog for audio.

Rhoads [21] describes a method that adds or subtracts small random quantities from each pixel. Addition or subtraction is determined by comparing a binary mask of L bits with the LSB of each pixel. If the LSB is equal to the corresponding mask bit, then the random quantity is added, otherwise it is subtracted. The watermark is subtracted by first computing the difference between the original and watermarked images and then by examining the sign of the difference, pixel by pixel, to determine if it corresponds to the original sequence of additions and subtractions. This method does not make use of perceptual relevance, but it is proposed that the high frequency noise be prefiltered to provide some robustness to lowpass filtering. This scheme does not consider the problem of collusion attacks.

Koch, Rindfrey, and Zhao [14] propose two general methods for watermarking images. The first method, attributed to Scott Burgett, breaks up an image into 8 × 8 blocks and computes the discrete cosine transform (DCT) of each of these blocks. A pseudorandom subset of the blocks is chosen, then, in each such block, a triple of frequencies is selected from one of 18 predetermined triples and modified so that their relative strengths encode a one or zero value. The 18 possible triples are composed by selection of three out of eight predetermined frequencies within the 8 × 8 DCT block. The choice of the eight frequencies to be altered within the DCT block is based on a belief that the "middle frequencies...have moderate variance," i.e. they have similar magnitude. This property is needed in order to allow the relative strength of the frequency triples to be altered without requiring a modification that would be perceptually noticeable. Superficially, this scheme is similar to our own proposal, also drawing an analogy to spread spectrum communications. However, the structure of their watermark is different from ours, and the set of frequencies is not chosen based on any direct perceptual significance, or relative energy considerations. Further, because the variance between the eight frequency coefficients is small, one would expect that their technique may be sensitive to noise or distortions. This is supported by the experimental results that report that the "embedded labels are robust against JPEG compression for a quality factor as low as about 50%." By comparison, we demonstrate that our method performs well with compression quality factors as low as 5%. An earlier proposal by Koch and Zhao [15] used not triples of frequencies but pairs of frequencies, and was again designed specifically for robustness to JPEG compression. Nevertheless, they state that "a lower quality factor will increase the likelihood that the changes necessary to superimpose the embedded code on the signal will be noticeably visible." In a second method, designed for black and white images, no frequency transform is employed. Instead, the selected blocks are modified so that the relative frequency of white and black pixels encodes the final value. Both watermarking procedures are particularly vulnerable to multiple document attacks. To protect against this, Zhao and Koch propose a *distributed* 8 × 8 block created by randomly sampling 64 pixels from the image. However, the resulting DCT has no relationship to that of the true image and consequently may be likely to cause noticeable artifacts in the image and be sensitive to noise.

In addition to direct work on watermarking images, there are several works of interest in related areas. Adelson [1] describes a technique for embedding digital information in an analog signal for the purpose of inserting digital data into an analog TV signal. The analog signal is quantized into one of two disjoint ranges ($\{0, 2, 4 \cdots\}$, $\{1, 3, 5 \cdots\}$, for example) that are selected based on the binary digit to be transmitted. Thus, Adelson's method is equivalent to watermark schemes that encode information into the LSB's of the data or its transform coefficients. Adelson recognizes that the method is susceptible to noise and therefore proposes an alternative scheme wherein a 2 × 1 Hadamard transform of the digitized analog signal is taken. The differential coefficient of the Hadamard transform is offset by zero or one unit prior to computing the inverse transform. This corresponds to encoding the watermark into the least significant bit of the differential coefficient of the Hadamard transform. It is not clear that this approach would demonstrate enhanced resilience to noise. Furthermore, like all such LSB schemes, an attacker can eliminate the watermark by randomization.

Schreiber *et al.* [22] describe a method to interleave a standard NTSC signal within an enhanced definition television (EDTV) signal. This is accomplished by analyzing the frequency spectrum of the EDTV signal (larger than that of the NTSC signal) and decomposing it into three subbands (L, M, H for low-, medium- and high-frequency, respectively). In contrast, the NTSC signal is decomposed into two subbands, L and M. The coefficients, M_k, within the M band are quantized into m levels and the high frequency coefficients, H_k, of the EDTV signal are scaled such that the addition of the H_k signal plus any noise present in the system is less than the minimum separation between quantization levels. Once more, the method relies on modifying least significant bits. Presumably, the midrange rather than low frequencies were chosen because these are less perceptually significant. In contrast, the method proposed here modifies the *most* perceptually significant components of the signal.

Finally, it should be noted that existing techniques are generally not resistant to collusion attacks by multiple documents.

III. Watermarking in the Frequency Domain

In order to understand the advantages of a frequency-based method, it is instructive to examine the processing stages that an image (or sound) may undergo in the process of copying, and to study the effect that these stages could have on the data, as illustrated in Fig. 1. In the figure, "transmission" refers to the application of any source or channel code, and/or standard encryption technique to the data. While most of these steps are information lossless, many compression schemes (JPEG, MPEG, etc.) are lossy, and can potentially degrade the data's quality, through *irretrievable* loss of information. In general, a watermarking scheme should be resilient to the distortions introduced by such algorithms.

Lossy compression is an operation that usually eliminates perceptually nonsalient components of an image or sound. Most processing of this sort takes place in the frequency domain. In fact, data loss usually occurs among the high-frequency components.

After receipt, an image may endure many common transformations that are broadly categorized as geometric distortions or signal distortions. Geometric distortions are specific to images and video, and include such operations as rotation, translation, scaling and cropping. By manually determining a minimum of four or nine corresponding points between the original and the distorted watermark, it is possible to remove any two or three-dimensional (3-D) affine transformation [8]. However, an affine scaling (shrinking) of the image leads to a loss of data in the high-frequency spectral regions of the image. Cropping, or the cutting out and removal of portions of an image, leads to irretrievable loss of image data, which may seriously degrade any spatially based watermark such as [6]. However, a frequency-based scheme spreads the watermark over the whole spatial extent of the image, and is therefore less likely to be affected by cropping, as demonstrated in Section V-E.

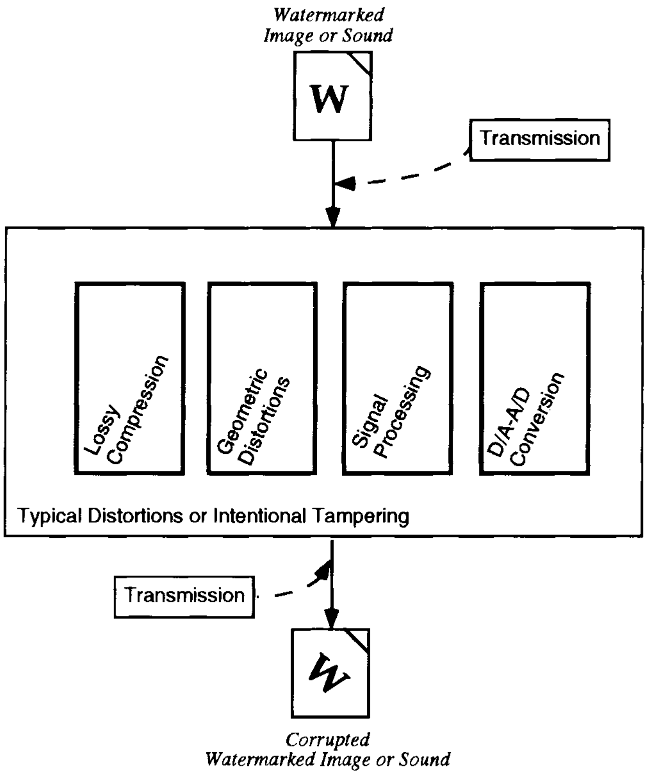


Fig. 1. Common processing operations that a media document could undergo.

Common signal distortions include digital-to-analog and analog-to-digital conversion, resampling, requantization, including dithering and recompression, and common signal enhancements to image contrast and/or color, and audio frequency equalization. Many of these distortions are nonlinear, and it is difficult to analyze their effect in either a spatial- or frequency-based method. However, the fact that the original image is known allows many signal transformations to be undone, at least approximately. For example, histogram equalization, a common nonlinear contrast enhancement method, may be removed substantially by histogram specification [10] or dynamic histogram warping [7] techniques.

Finally, the copied image may not remain in digital form. Instead, it is likely to be printed, or an analog recording made (onto analog audio or video tape). These reproductions introduce additional degradation into the image that a watermarking scheme must be robust to.

The watermark must not only be resistant to the inadvertent application of the aforementioned distortions. It must also be immune to intentional manipulation by malicious parties. These manipulations can include combinations of the above distortions, and can also include collusion and forgery attacks, which are discussed in Section IV-E.

A. Spread Spectrum Coding of a Watermark

The above discussion illustrates that the watermark should *not* be placed in perceptually insignificant regions of the image (or its spectrum), since many common signal and geometric processes affect these components. For example, a watermark placed in the high-frequency spectrum of an image can be easily eliminated with little degradation to the image by any process that directly or indirectly performs lowpass filtering. The problem then becomes how to insert a watermark into the most perceptually significant regions of the spectrum in a fidelity preserving fashion. Clearly, any spectral coefficient may be altered, provided such modification is small. However, very small changes are very susceptible to noise.

To solve this problem, the frequency domain of the image or sound at hand is viewed as a *communication channel*, and correspondingly, the watermark is viewed as a signal that is transmitted through it. Attacks and unintentional signal distortions are thus treated as noise that the immersed signal must be immune to. While we use this methodology to hide watermarks in data, the same rationale can be applied to sending any type of message through media data.

We originally conceived our approach by analogy to spread spectrum communications [20]. In spread spectrum communications, one transmits a narrowband signal over a much larger bandwidth such that the signal energy present in any single frequency is undetectable. Similarly, the watermark is spread over very many frequency bins so that the energy in any one bin is very small and certainly undetectable. Nevertheless, because the watermark verification process knows the location and content of the watermark, it is possible to concentrate these many weak signals into a single output with high signal-to-noise ratio (SNR). However, to destroy such a watermark would require noise of high amplitude to be added to *all* frequency bins.

Spreading the watermark throughout the spectrum of an image ensures a large measure of security against unintentional or intentional attack: First, the location of the watermark is not obvious. Furthermore, frequency regions should be selected in a fashion that ensures severe degradation of the original data following any attack on the watermark.

A watermark that is well placed in the frequency domain of an image or a sound track will be practically impossible to see or hear. This will always be the case if the energy in the watermark is sufficiently small in any single frequency coefficient. Moreover, it is possible to increase the energy present in particular frequencies by exploiting knowledge of masking phenomena in the human auditory and visual systems. Perceptual masking refers to any situation where information in certain regions of an image or a sound is occluded by perceptually more prominent information in another part of the scene. In digital waveform coding, this frequency domain (and, in some cases, time/pixel domain) masking is exploited extensively to achieve low bit rate encoding of data [9], [12]. It is known that both the auditory and visual systems attach more resolution to the high-energy, low-frequency, spectral regions of an auditory or visual scene [12]. Further, spectrum analysis of images and sounds reveals that most of the information in such data is located in the low-frequency regions.

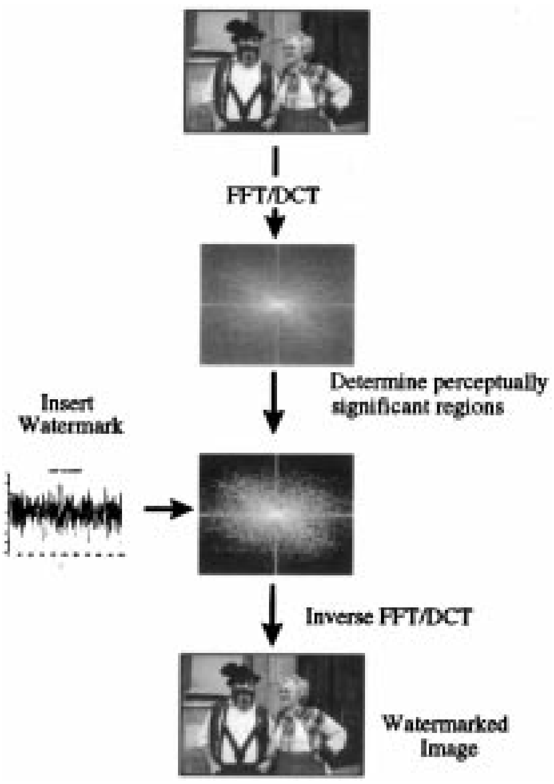


Fig. 2. Stages of watermark insertion process.

Fig. 2 illustrates the general procedure for frequency domain watermarking. Upon applying a frequency transformation to the data, a *perceptual mask* is computed that highlights perceptually significant regions in the spectrum that can support the watermark without affecting perceptual fidelity. The watermark signal is then inserted into these regions in a manner described in Section IV-B. The precise magnitude of each modification is only known to the owner. By contrast, an attacker may only have knowledge of the possible range of modification. To be confident of eliminating a watermark, an attacker must assume that each modification was at the limit of this range, despite the fact that few such modifications are typically this large. As a result, an attack creates visible (or audible) defects in the data. Similarly, unintentional signal distortions due to compression or image manipulation, must leave the perceptually significant spectral components intact, otherwise the resulting image will be severely degraded. This is why the watermark is robust.

In principle, any frequency domain transform can be used. However, in the experimental results of Section VI we use a Fourier domain method based on the DCT [16], although we are currently exploring the use of wavelet-based schemes as a variation. In our view, each coefficient in the frequency domain has a *perceptual capacity*, that is, a quantity of additional information can be added without any (or with minimal) impact to the perceptual fidelity of the data. To determine the perceptual capacity of each frequency, one can use models for the appropriate perceptual system or simple experimentation.

In practice, in order to place a length n watermark into an $N \times N$ image, we computed the $N \times N$ DCT of the image and placed the watermark into the n highest magnitude coefficients of the transform matrix, excluding the DC component.[1] For most images, these coefficients will be the ones corresponding to the low frequencies.

In the next section, we provide a high level discussion of the watermarking procedure, describing the structure of the watermark and its characteristics.

IV. Structure of the Watermark

We now give a high-level overview of our a basic watermarking scheme; many variations are possible. In its most basic implementation, a watermark consists of a sequence of real numbers $X = x_1, \cdots, x_n$. In practice, we create a watermark where each value x_i is chosen independently according to $N(0,1)$ (where $N(\mu, \sigma^2)$ denotes a normal distribution with mean μ and variance σ^2). We assume that numbers are represented by a reasonable but finite precision and ignore these insignificant roundoff errors. Section IV-A introduces notation to describe the insertion and extraction of a watermark and Section IV-D describes how two watermarks (the original one and the recovered, possibly corrupted one) can be compared. This procedure exploits the fact that each component of the watermark is chosen from a normal distribution. Alternative distributions are possible, including choosing x_i uniformly from $\{1, -1\}, \{0, 1\}$ or $[0, 1]$. However, as we discuss in IV-D, using such distributions leaves one particularly vulnerable to attacks using multiple watermarked documents.

A. Description of the Watermarking Procedure

We extract from each document D a sequence of values $V = v_1, \cdots, v_n$, into which we insert a watermark $X = x_1, \cdots, x_n$ to obtain an adjusted sequence of values $V' = v'_1, \cdots, v'_n$. V' is then inserted back into the document in place of V to obtain a watermarked document D'. One or more attackers may then alter D', producing a new document D^*. Given D and D^*, a possibly corrupted watermark X^* is extracted and is compared to X for statistical significance. We extract X^* by first extracting a set of values $V^* = v^*_1, \ldots, v^*_n$ from D^* (using information about D) and then generating X^* from V^* and V.

Frequency-domain based methods for extracting V and V^* and inserting V' are given in Section III. For the rest of this section, we ignore the manipulations of the underlying documents.

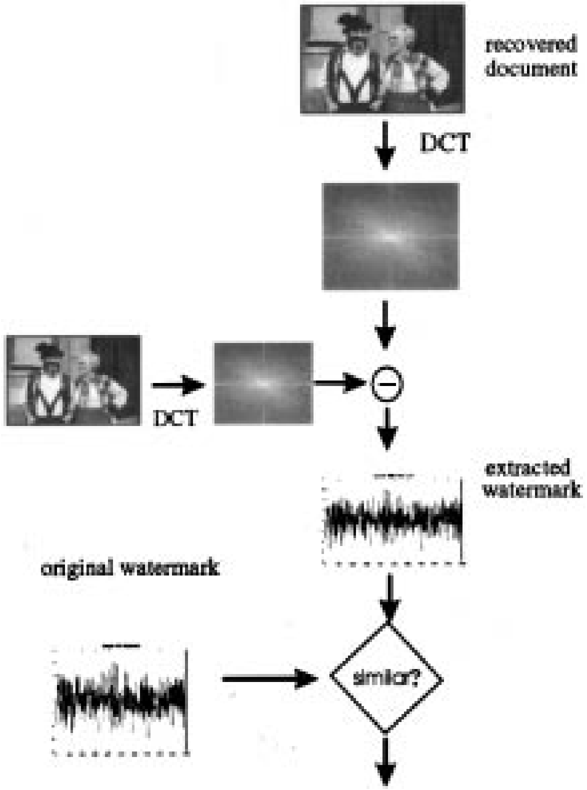


Fig. 3. Encoding and decoding of the watermark string.

B. Inserting and Extracting the Watermark

When we insert X into V to obtain V' we specify a scaling parameter α, which determines the extent to which X alters V. Three natural formulae for computing V' are

$$v'_i = v_i + \alpha x_i \quad (1)$$
$$v'_i = v_i(1 + \alpha x_i) \quad (2)$$
$$v'_i = v_i(e^{\alpha x_i}). \quad (3)$$

Equation (1) is always invertible, and (2) and (3) are invertible if $v_i \neq 0$, which holds in all of our experiments. Given V^*, we can therefore compute the inverse function to derive X^* from V^* and V.

Equation (1) may not be appropriate when the v_i values vary widely. If $v_i = 10^6$ then adding 100 may be insufficient for establishing a mark, but if $v_i = 10$ adding 100 will distort this value unacceptably. Insertion based on (2) or (3) are more robust against such differences in scale. We note that (2) and (3) give similar results when αx_i is small. Also, when v_i is positive, then (3) is equivalent to $\lg(v'_i) = \lg(v_i) + \alpha x_i$, and may be viewed as an application of (1) to the case where the logarithms of the original values are used.

1) Determining Multiple Scaling Parameters: A single scaling parameter α may not be applicable for perturbing all of the values v_i, since different spectral components may exhibit more or less tolerance to modification. More generally one can have multiple scaling parameters $\alpha_1, \cdots, \alpha_n$ and use update rules such as $v'_i = v_i(1 + \alpha_i x_i)$. We can view α_i as a relative measure of how much one must alter v_i to alter the perceptual quality of the document. A large α_i means that one

[1] More generally, n randomly chosen coefficients could be chosen from the M, $M \geq n$ most perceptually significant coefficients of the transform. The choice of appropriate components remains a subject of research.

can perceptually "get away" with altering v_i by a large factor without degrading the document.

There remains the problem of selecting the multiple scaling values. In some cases, the choice of α_i may be based on some general assumption. For example, (2) is a special case of the generalized (1) $(v_i' = v_i + \alpha_i x_i)$, for $\alpha_i = \alpha v_i$. Essentially, (2) makes the reasonable assumption that a large value is less sensitive to additive alterations than a small value.

In general, one may have little idea of how sensitive the image is to various values. One way of empirically estimating these sensitivities is to determine the distortion caused by a number of attacks on the original image. For example, one might compute a degraded image D^* from D, extract the corresponding values $v_1^*, \cdots, v_n^*$ and choose α_i to be proportional to the deviation $|v_i^* - v_i|$. For greater robustness, one should try many forms of distortion and make α_i proportional to the average value of $|v_i^* - v_i|$. As alternatives to taking the average deviation one might also take the median or maximum deviation.

One may combine this empirical approach with general global assumptions about the sensitivity of the values. For example, one might require that $\alpha_i \geq \alpha_j$ whenever $v_i \geq v_j$. One way to combine this constraint with the empirical approach would be to set α_i according to

$$\alpha_i \sim \max_{j|v_j \leq v_i} |v_j^* - v_j|.$$

A still more sophisticated approach would be to weaken the monotonicity constraint to be robust against occasional outliers.

In all our experiments we simply use (2) with a single parameter $\alpha = 0.1$. When we computed JPEG-based distortions of the original image, we observed that the higher energy frequency components were not altered proportional to their magnitude [the implicit assumption of (2)]. We suspect that we could make a less obtrusive mark of equal strength by attenuating our alterations of the high-energy components and amplifying our alterations of the lower energy components. However, we have not yet performed this experiment.

C. Choosing the Length, n, of the Watermark

The choice of n dictates the degree to which the watermark is spread out among the relevant components of the image. In general, as the number of altered components are increased the extent to which they must be altered decreases. For a more quantitative assessment of this tradeoff, we consider watermarks of the form $v_i' = v_i + \alpha x_i$ and model a white noise attack by $v_i^* = v_i' + r_i$ where r_i are chosen according to independent normal distributions with standard deviation σ. For the watermarking procedure we described below, one can recover the watermark when α is proportional to $\sigma/\sqrt{n}$. That is, by quadrupling the number of components used, one can halve the magnitude of the watermark placed into each component. Note that the sum of squares of the deviations will be essentially unchanged.

Note that the number of bits of information associated with the watermark can be arbitrary—the watermark is simply used as an index to a database entry associated with the watermark.

D. Evaluating the Similarity of Watermarks

It is highly unlikely that the extracted mark X^* will be identical to the original watermark X. Even the act of requantizing the watermarked document for delivery will cause X^* to deviate from X. We measure the similarity of X and X^* by

$$\mathrm{sim}(X, X^*) = \frac{X^* \cdot X}{\sqrt{X^* \cdot X^*}}. \qquad (4)$$

Many other measures are possible, including the standard correlation coefficient. Further variations on this basic metric are discussed in IV-D2. To decide whether X and X^* match, one determines whether $\mathrm{sim}(X, X^*) > T$, where T is some threshold. Setting the detection threshold is a classical decision estimation problem in which we wish to minimize both the rate of false negatives (missed detections) and false positives (false alarms) [23]. We have chosen our measure so that it is particularly easy to determine the probability of false positives.

1) Computing the Probability of False Positives: There is always the possibility that X and X^* will be very similar purely by random chance; hence, any similarity metric will give "significant" values that are spurious. We analyze the probability of such false positives as follows. Suppose that the creators of document D^* had no access to X (either through the seller or through a watermarked document). Then, even conditioned on any fixed value for X^*, each x_i will be independently distributed according to $N(0,1)$. That is, X is independent of X^*.

The distribution on $X^* \cdot X$ may be computed by first writing it as $\sum_{i=1}^n x_i^* x_i$, where x_i^* is a constant. Using the well-known formula for the distribution of a linear combination of variables that are independent and normally distributed, $X^* \cdot X$ will be distributed according to

$$N\left(0, \sum_{i=1}^{n} {x_i^*}^2\right) = N(0, X^* \cdot X^*),$$

Thus, $\mathrm{sim}(X, X^*)$ is distributed according to $N(0,1)$. We can then apply the standard significance tests for the normal distribution. For example, if X^* is created independently from X then the probability that $\mathrm{sim}(X, X^*) > 6$ is the probability of a normally distributed random variable exceeding its mean by more than six standard deviations.

Hence, for a small number of documents, setting the threshold at T equal to six will cause spurious matchings to be extremely rare. Of course, the number of tests to be performed must be considered in determining what false positive probability is acceptable. For example, if one tests an extracted watermark X^* against 10^6 watermarks, then the probability of a false positive is increased by a multiplicative factor of 10^6 as well.

We note that our similarity measure and the false-positive probability analysis does not depend on n, the size of the watermark. However, n implicitly appears, since for example, $\mathrm{sim}(X, X)$ is likely to be around $\sqrt{n}$ when X is generated in the prescribed manner. As a rule of thumb, larger values of n tend to cause larger similarity values when X and X^* are genuinely related (e.g., X^* is a distorted version of X),

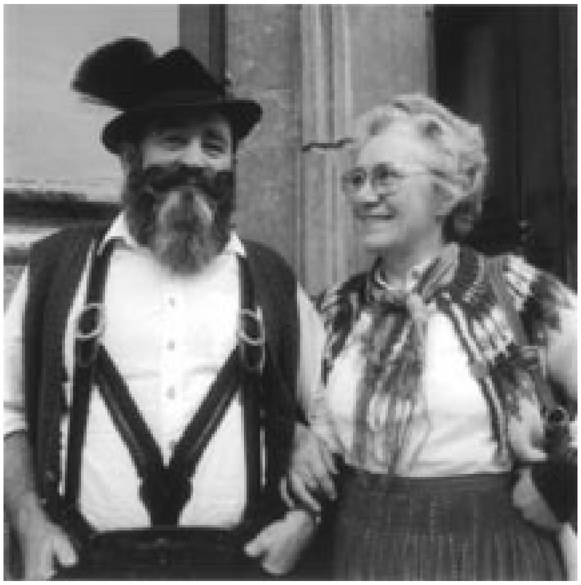


Fig. 4. Bavarian couple image courtesy of Corel Stock Photo Library.

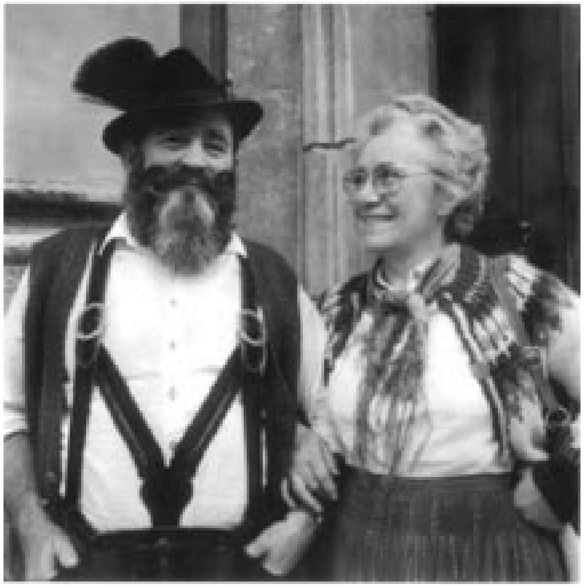


Fig. 5. Watermarked version of Bavarian couple.

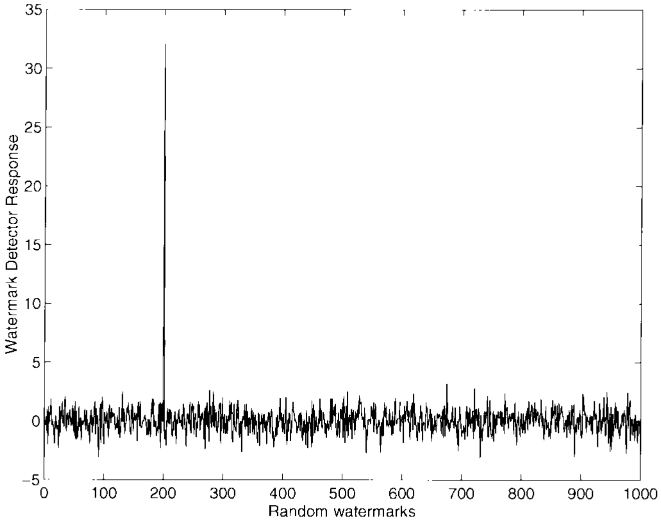


Fig. 6. Watermark detector response to 1000 randomly generated watermarks. Only one watermark (the one to which the detector was set to respond) matches that present in Fig. 5.

without causing larger similarity values when X and X^* are independent. This benefit must be balanced against the tendency for the document to be more distorted when n is larger.

a) A remark on quantization: In the above analysis, we treated all of the vectors as consisting of ideal real numbers. In practice, the actual values inserted will be quantized to some extent. Nevertheless, it is simpler to view the watermarks as real numbers and the quantization process as yet another form of distortion. Our analysis of false positives does not depend on the distribution or even the domain of possible X^*, and hence holds regardless of quantization effects.

There is an additional, extremely low-order quantization effect that occurs because X is generated with only finite precisions. However, this effect is caused only by the arithmetic precision, and not on the constraints imposed by the document. If each $x_i \in X$ is stored as a double-precision real number, the difference between the calculated value of $\mathsf{sim}(X, X^*)$ and its "ideal" value will be quite small for any reasonable n and any reasonable bound on the dynamic range of X^*.

2) Robust Statistics The above analysis required only the independence of X from X^*, and did not rely on any specific properties of X^* itself. This fact gives us further flexibility when it comes to preprocessing X^*. We can process X^* in a number of ways to potentially enhance our ability to extract a watermark. For example, in our experiments on images we encountered instances where the average value of x_i^*, denoted $E_i(X^*)$, differed substantially from zero, due to the effects of a dithering procedure. While this artifact could be easily eliminated as part of the extraction process, it provides a motivation for postprocessing extracted watermarks. We found that the simple transformation $x_i^* \leftarrow x_i^* - E_i(X^*)$ yielded superior values of $\mathsf{sim}(X, X^*)$. The improved performance resulted from the decreased value of $X^* \cdot X^*$; the value of $X^* \cdot X$ was only slightly affected.

In our experiments, we frequently observed that x_i^* could be greatly distorted for some values of i. One postprocessing option is to simply ignore such values, setting them to zero. That is

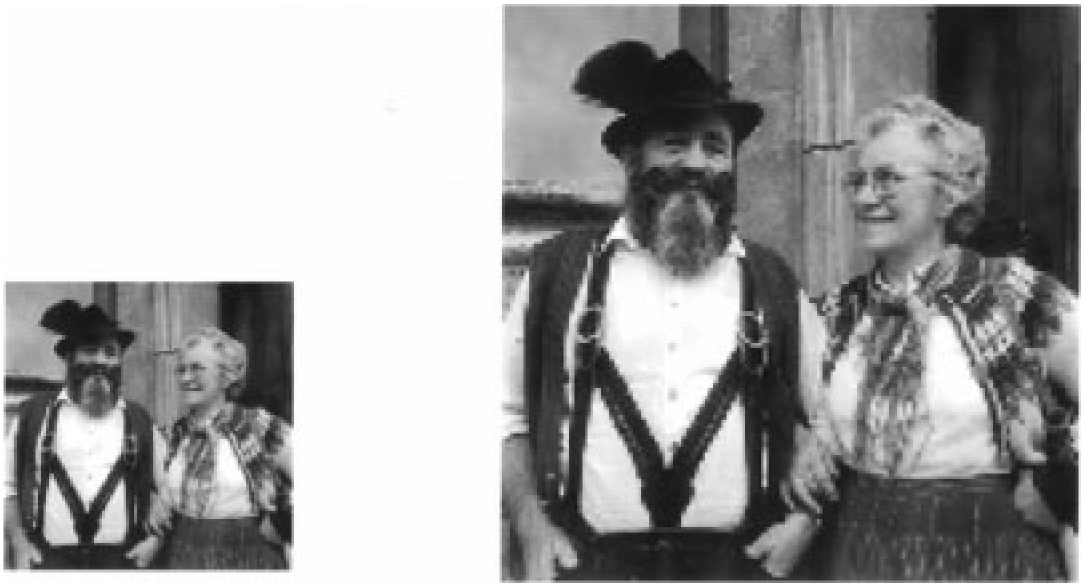
$$x_i^* \leftarrow \begin{cases} x_i^*, & \text{if } |x_i^*| \le \text{tolerance} \\ 0, & \text{otherwise.} \end{cases}$$

Again, the goal of such a transformation is to lower $X^* \cdot X^*$. A less abrupt version of this approach is to normalize the X^* values to be either $-1, 0$ or 1, by

$$x_i^* \leftarrow \mathsf{sign}(x_i^* - E_i(X^*)).$$

This transformation can have a dramatic effect on the statistical significance of the result. Other robust statistical techniques could also be used to suppress outlier effects [11].

A natural question is whether such postprocessing steps run the risk of generating false positives. Indeed, the same potential risk occurs whenever there is any latitude in the



(a) (b)

Fig. 7. (a) Lowpass filtered, 0.5 scaled image of Bavarian couple. (b) Rescaled image showing noticeable loss of fine detail.

procedure for extracting X^* from D^*. However, as long as the method for generating a set of values for X^* depends solely on D and D^*, our statistical significance calculation is unaffected. The only caveat to be considered is that the bound on the probability that one of $X_1^*, \cdots X_k^*$ generates a false positive is the sum of the individual bounds. Hence, to convince someone that a watermark is valid, it is necessary to have a published and rigid extraction and processing policy that is guaranteed to only generate a small number of candidate X^*.

E. Resilience to Multiple-Document (Collusion) Attacks

The most general attack consists of using t multiple watermarked copies $D_1', \cdots, D_t'$ of document D to produce an unwatermarked document D^*. We note that most schemes proposed seem quite vulnerable to such attacks. As a theoretical exception, Boneh and Shaw [5] propose a coding scheme for use in situations in which one can insert many relatively weak $0/1$ watermarks into a document. They assume that if the ith watermark is the same for all t copies of the document then it cannot be detected, changed or removed. Using their coding scheme, the number of weak watermarks to be inserted scales according to t^4, which may limit its usefulness in practice.

To illustrate the power of multiple-document attacks, consider watermarking schemes in which v_i' is generated by either adding 1 or -1 at random to v_i. Then as soon as one finds two documents with unequal values for v_i', one can determine v_i and, hence, completely eliminate this component of the watermark. With t documents one can, on average, eliminate all but a 2^{1-t} fraction of the components of the watermark. Note that this attack does not assume anything about the distribution on v_i. While a more intelligent allocation of ± 1 values to the watermarks (following [5] and [17]) will better resist this simple attack, the discrete nature of the watermark components makes them much easier to completely eliminate. Our use of continuous valued watermarks appears to give greater resilience to such attacks. Interestingly, we have experimentally determined that if one chooses the x_i uniformly over some range, then one can remove the watermark using only five documents.

Use of the normal distribution seems to give better performance than the distributions considered above. We note that the crucial performance measure to consider is the value of $\max_i(X^* \cdot X_i)$, where X^* is the watermark extracted from an document D^* generated by attacking documents $D_1, \cdots, D_t$, with respective watermarks $X_1, \cdots, X_t$. The denominator $\sqrt{X^* \cdot X^*}$ of our similarity measure can always be made larger by, for example, adding noise. This causes the similarity measure to shrink, at the expense of distorting the image. Hence, we can view $\max_i(X^* \cdot X_i)$ as determining a fidelity/undetectability tradeoff curve and the value of $\sqrt{X^* \cdot X^*}$ as picking a point on this curve.

When X_i is inserted into D by a linear update rule, then an averaging attack, which sets

$$D^* = \frac{D_1 + \cdots + D_t}{t}$$

will result in

$$X^* = \frac{X_1 + \cdots + X_t}{t}.$$

In this case,

$$\max_i(X^* \cdot X_i) \approx \frac{1}{t} \max_i(X_i \cdot X_i) \text{ (assuming } X_i X_j \approx 0).$$

That is, there is a $1/t$ behavior in the detector output.

Note that with a naive averaging attack, the denominator, $\sqrt{X^* \cdot X^*}$, will be a (roughly) $1/\sqrt{t}$ factor smaller, so $\max_i \mathsf{sim}(X_i, X^*)$ will be roughly $\sqrt{n}/\sqrt{t}$. However, as mentioned before, additional noise can be added so that the extracted watermark, X^*, has the same power as any of the original watermarks X_i. Then $\max_i \mathsf{sim}(X_i, X^*)$ will be

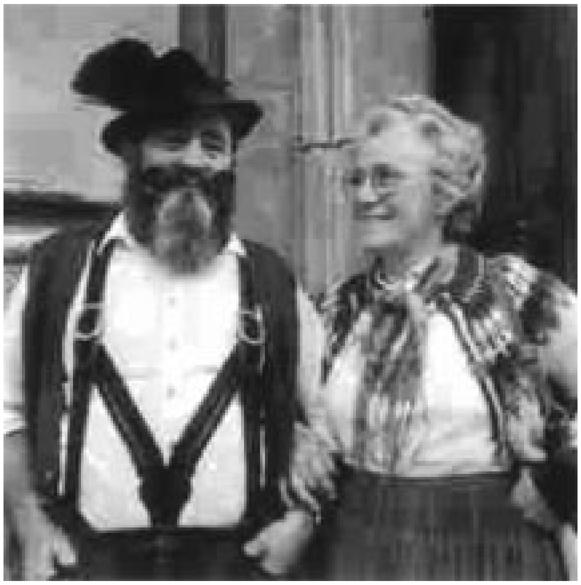


Fig. 8. JPEG encoded version of Bavarian couple with 10% quality and 0% smoothing.

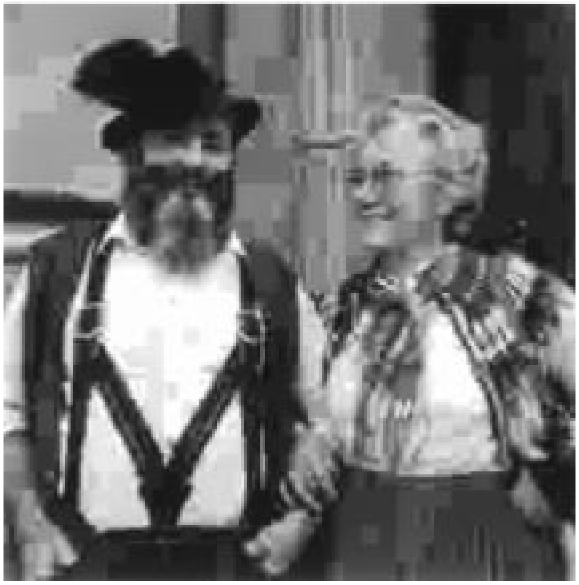


Fig. 9. JPEG encoded version of Bavarian couple with 5% quality and 0% smoothing.

roughly $\sqrt{n}/t$. Thus, the similarity measure can be shrunk by a factor of t.

We do not know of any more effective multidocument attack on normally distributed watermarks. In a forthcoming paper (see http://www.neci.nj.nec.com/tr/index.html), a more theoretical justification is given for why it is hard to achieve more than an $O(t)$ reduction in the similarity measure.

V. Experimental Results

In order to evaluate the proposed watermarking scheme, we took the Bavarian couple[2] image of Fig. 4 and produced the watermarked version of Fig. 5. We then subjected the watermarked image to a series of image processing and collusion style attacks. These experiments are preliminary, but show resilience to certain types of common processing. Of note is our method's resistance to compression such as JPEG, and data conversion (printing, xeroxing and scanning). Note that in the case of affine transforms, registration to the original image is crucial to successful extraction.

In all experiments, a watermark length of 1000 was used. We added the watermark to the image by modifying 1000 of the more perceptually significant components of the image spectrum using (2). More specifically, the 1000 largest coefficients of the DCT (excluding the DC term) were used. A fixed scale factor of 0.1 was used throughout.

A. Experiment 1: Uniqueness of Watermark

Fig. 6 shows the response of the watermark detector to 1000 randomly generated watermarks of which only one matches the watermark present in Fig. 5. The positive response due to the correct watermark is very much stronger that the response to incorrect watermarks, suggesting that the algorithm has very low false positive response rates.

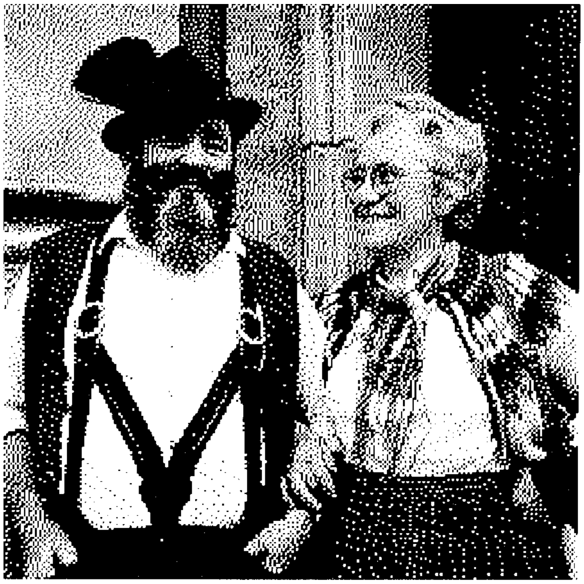


Fig. 10. Dithered version of the Bavarian couple image.

B. Experiment 2: Image Scaling

We scaled the watermarked image to half of its original size, as shown in Fig. 7(a). In order to recover the watermark, the quarter-sized image was rescaled to its original dimensions, as shown in Fig. 7(b), in which it is clear that considerable fine detail has been lost in the scaling process. This is to be expected since subsampling of the image requires a lowpass spatial filtering operation. The response of the watermark detector to the original watermarked image of Fig. 5 was 32.0, which compares to a response of 13.4 for the rescaled version of Fig. 7(b). While the detector response is down by over 50%, the response is still well above random chance

[2] The common test image Lenna was originally used in our experiments, and similar results were obtained. However, Playboy Inc. refused to grant copyright permission for electronic distribution.

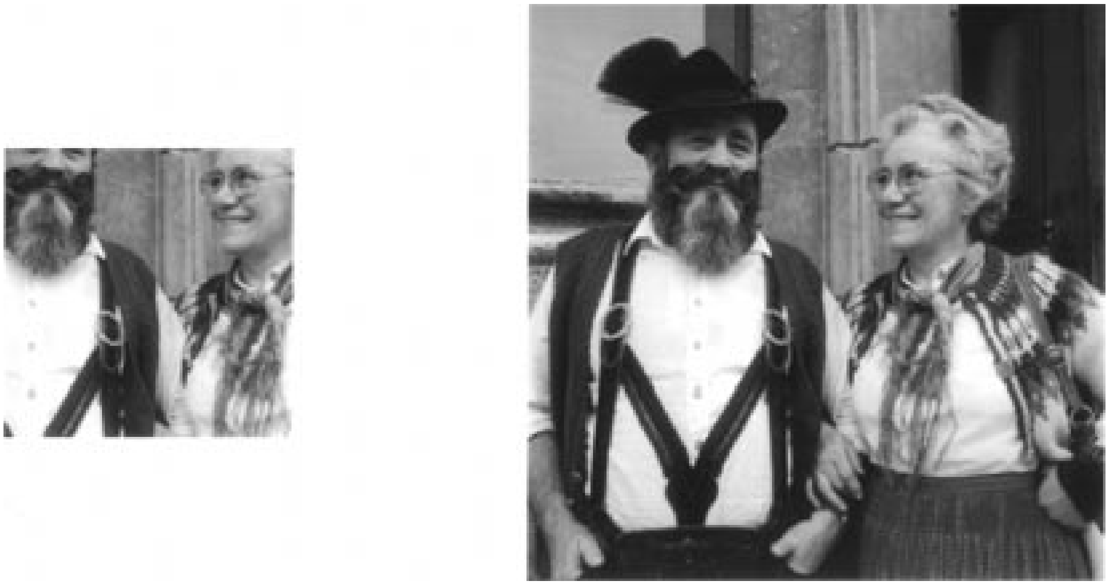


Fig. 11. (a) Clipped version of watermarked Bavarian couple. (b) Restored version of Bavarian couple in which missing portions have been replaced with imagery from the original unwatermarked image of Fig. 4.

levels suggesting that the watermark is robust to geometric distortions. Moreover, it should be noted that 75% of the original data is missing from the scaled down image of Fig. 7.[3]

C. Experiment 3: JPEG Coding Distortion

Fig. 8 shows a JPEG encoded version of the Bavarian couple image with parameters of 10% quality and 0% smoothing, which results in clearly visible distortions of the image. The response of the watermark detector is 22.8, again suggesting that the algorithm is robust to common encoding distortions. Fig. 9 shows a JPEG encoded version of Bavarian couple with parameters of 5% quality and 0% smoothing, which results is very significant distortions of the image. The response of the watermark detector in this case is 13.9, which is still well above random.

D. Experiment 4: Dithering Distortion

Fig. 10 shows a dithered version of Bavarian couple. The response of the watermark detector is 5.2, again suggesting that the algorithm is robust to common encoding distortions. In fact, more reliable detection can be achieved simply by removing any nonzero mean from the extracted watermark, as discussed in Section IV-D2. In this case the detection value is 10.5.

E. Experiment 5: Cropping

Fig. 11(a) shows a cropped version of the watermarked image of Fig. 5 in which only the central quarter of the image remains. In order to extract the watermark from this image, the missing portions of the image were replaced with portions from the original *unwatermarked* image of Fig. 4, as shown in Fig. 11(b). In this case, the response of the watermark is 14.6. Once again, this is well above random even though 75% of the data has been removed.

Fig. 12(a) shows a clipped version of the JPEG encoded image of Fig. 8 in which only the central quarter of the image remains. As before, the missing portions of the image were replaced with portions from the original *unwatermarked* image of Fig. 4, as shown in Fig. 12(b). In this case, the response of the watermark is 10.6. Once more, this is well above random even though 75% of the data has been removed and distortion is present in the clipped portion of the image.

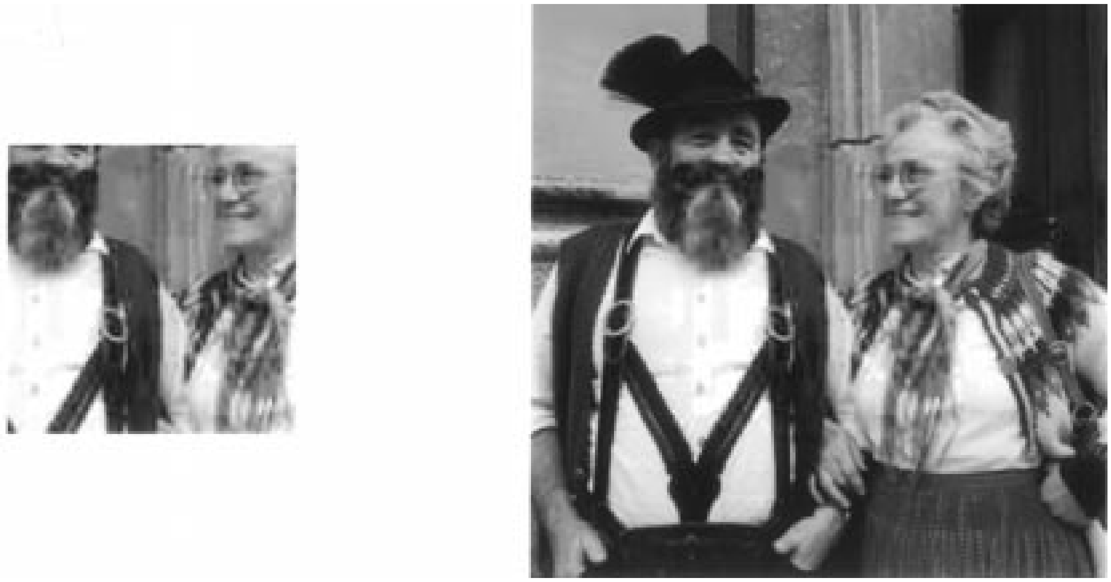
F. Experiment 6: Print, Xerox, and Scan

Fig. 13 shows an image of the Bavarian Couple after 1) printing, 2) xeroxing, then 3) scanning at 300 dpi using a UMAX PS-2400X scanner, and finally 4) rescaling to a size of 256 × 256. Clearly, this image suffers from several levels of distortion that accompany each of the four stages. High-frequency pattern noise is especially noticeable. The detector response to the watermark is 4.0. However, if the nonzero mean is removed and only the sign of the elements of the watermark are used, then the detector response is 7.0, which is well above random.

G. Experiment 7: Attack by Watermarking Watermarked Images

Fig. 14 shows an image of Bavarian Couple after five successive watermarking operations, i.e., the original image is watermarked, the watermarked image is watermarked, etc. This may be considered another form of attack in which it is clear that significant image degradation eventually occurs as the process is repeated. This attack is equivalent to adding noise to the frequency bins containing the watermark. Interestingly, Fig. 15 shows the response of the detector to

[3] However, subsequent experiments have revealed that if small changes of scale are not corrected, then the response of the watermark detector is severely degraded.



(a) (b)

Fig. 12. (a) Clipped version of JPEG encoded (10% quality, 0% smoothing) Bavarian couple. (b) Restored version of Bavarian couple in which missing portions have been replaced with imagery from the original unwatermarked image of Fig. 4.

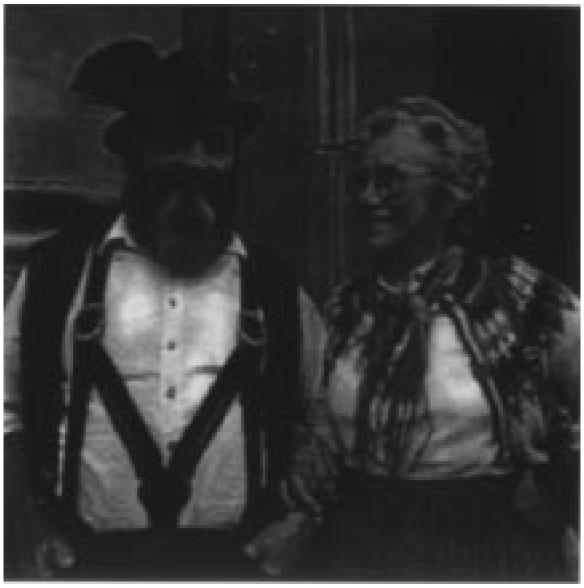


Fig. 13. Printed, xeroxed, scanned, and rescaled image of Bavarian couple.

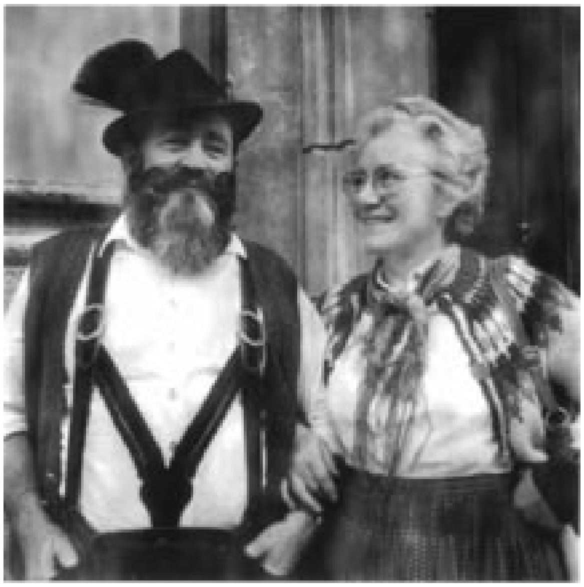


Fig. 14. Image of Bavarian couple after five successive watermarks have been added.

1000 randomly generated watermarks, which include the five watermarks present in the image. Five spikes clearly indicate the presence of the five watermarks and demonstrate that successive watermarking does not unduly interfere with the process.

H. Experiment 8: Attack by Collusion

In a similar experiment, we took five separately watermarked images and averaged them to form Fig. 16 in order to simulate a simple collusion attack. As before, Fig. 17 shows the response of the detector to 1000 randomly generated watermarks, which include the five watermarks present in the image. Once again, five spikes clearly indicate the presence of the five watermarks and demonstrate that simple collusion based on averaging a few images is an ineffective attack.

VI. Conclusion

A need for electronic watermarking is developing as electronic distribution of copyright material becomes more prevalent. Above, we outlined the necessary characteristics of such a watermark. These are: fidelity preservation, robustness to common signal and geometric processing operations, robustness to attack, and applicability to audio, image and video data.

To meet these requirements, we propose a watermark whose structure consists of k i.i.d. random numbers drawn from a $N(0,1)$ distribution. We rejected a binary watermark because it is far less robust to attacks based on collusion of several independently watermarked copies of an image. The length of the watermark is variable and can be adjusted to suit the characteristics of the data. For example, longer watermarks

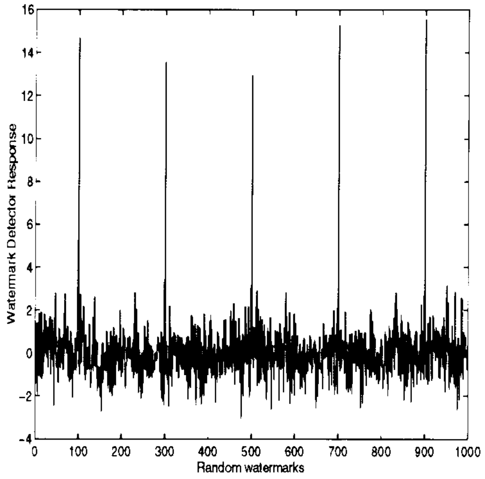


Fig. 15. Watermark detector response to 1000 randomly generated watermarks (including the five specific watermarks) for the watermarked image of Fig. 14. Each of the five watermarks is clearly indicated.

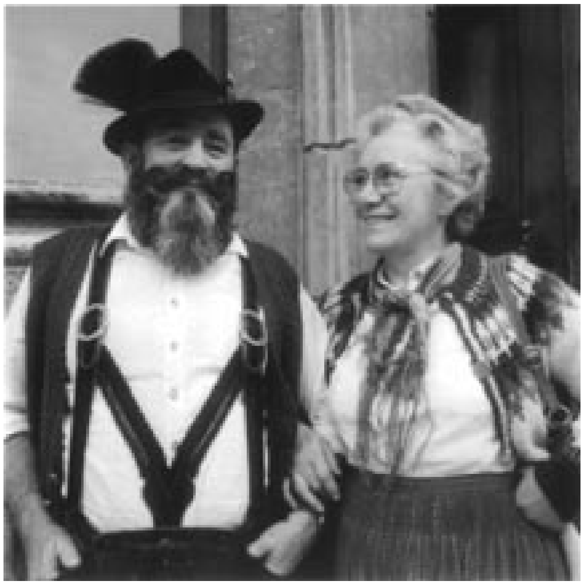


Fig. 16. Image of Bavarian couple after averaging together five independently watermarks versions of the Bavarian couple image.

may be used for an image that is especially sensitive to large modifications of its spectral coefficients, thus requiring weaker scaling factors for individual components.

We recommend that the watermark be placed in the perceptually *most* significant components of the image spectrum. This maximizes the chances of detecting the watermark even after common signal and geometric distortions. Further, modification of these spectral components results in severe image degradation long before the watermark itself is destroyed. Of course, to insert the watermark, it is necessary to alter these very same coefficients. However, each modification can be extremely small and, in a manner similar to spread spectrum communication, a strong narrowband watermark may be distributed over a much broader image (channel) spectrum. We have not performed an objective evaluation of the image quality, in part because the image quality can be adjusted to any desired quality by altering the relative power of the watermark using the scale factor term. Of course, as the watermark strength is reduced to improve the image quality, the robustness of the method is also reduced. It will ultimately be up to content owners to decide what image degradation and what level of robustness is acceptable. This will vary considerably from application to application.

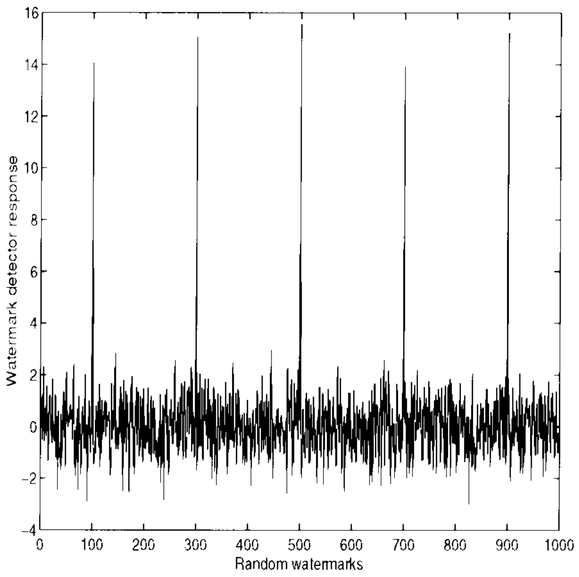


Fig. 17. Watermark detector response to 1000 randomly generated watermarks (including the five specific watermarks) for the watermarked image of Fig. 16. Each of the five watermarks is clearly detected, indicating that collusion by averaging is ineffective.

Detection of the watermark then proceeds by adding all of these very small signals, and concentrating them once more into a signal with high SNR. Because the magnitude of the watermark at each location is only known to the copyright holder, an attacker would have to add much more noise energy to each spectral coefficient in order to be sufficiently confident of removing the watermark. However, this process would destroy the image fidelity.

In our experiments, we added the watermark to the image by modifying the 1000 largest coefficients of the DCT (excluding the DC term). These components are heuristically perceptually more significant than others. An important open problem is the construction of a method that would identify perceptually significant components from an analysis of the image and the human perceptual system. Such a method may include additional considerations regarding the relative predictability of a frequency based on its neighbors. The latter property is important in combating attacks that may use statistical analyzes of frequency spectra to replace components with their maximum likelihood estimate. For example, the choice of the DCT is not critical to the algorithm and other spectral transforms, including wavelet type decompositions, are also possible.

We showed, using the Bavarian couple image, that our algorithm can extract a reliable copy of the watermark from imagery that we degraded with several common geometric and signal processing procedures. An important caveat here is that any affine geometric transformation must first be inverted. These procedures include translation, rotation, scale change, and cropping. The algorithm displays strong resilience to lossy operations such as aggressive scale changes, JPEG compression, dithering and data conversion. The experiments presented are preliminary, and should be expanded in order to validate the results. We are conducting ongoing work in this area. Further, the degree of precision of the registration procedures used in undoing affine transforms must be characterized precisely across a large test set of images.

Application of the method to color images is straightforward. The most common transformation of a color image is to convert it to black and white. Color images are therefore converted into a YIQ representation and the brightness component Y is then watermarked. The color image can then be converted to other formats, but must be converted back to YIQ prior to extraction of the watermark. We therefore expect color images to be robust to the signal transformations we applied to gray-level images. However, robustness to certain color image processing procedures should be investigated. Similarly, the system should work well on text images, however, the binary nature of the image together with its much more structured spectral distribution need more work. We expect that our watermarking methodology should extend straightforwardly to audio and video data. However, special attention must be paid to the time-varying nature of these data.

Broader systems issues must be also addressed in order for this system to be used in practice. For example, it would be useful to be able to prove in court that a watermark is present without publicly revealing the original, unmarked document. This is not hard to accomplish using secure trusted hardware; an efficient purely cryptographic solution seems much more difficult. It should also be noted that the current proposal only allows the watermark to be extracted by the owner, since the original unwatermarked image is needed as part of the extraction process. This prohibits potential users from querying the image for ownership and copyright information. This capability may be desirable but appears difficult to achieve with the same level of tamper resistance. However, it is straightforward to provide if a much weaker level of protection is acceptable and might therefore be added as a secondary watermarking procedure. Finally, we note that while the proposed methodology is used to hide watermarks in data, the same process can be applied to sending other forms of message through media data.

Acknowledgment

I. Cox and T. Shamoon thank L. O'Gorman of AT&T Bell Laboratories for bringing this problem to their attention, and S. Roy for testing the robustness of the algorithm. I. Cox thanks H. Stone for advice on image transforms.

References

[1] E. H. Adelson, "Digital signal encoding and decoding apparatus," U.S. Patent 4 939 515, 1990.

[2] G. W. Braudaway, K. A. Magerlein, and F. C. Mintzer, "Color correct digital watermarking of images," U.S. Patent 5 530 759, 1996.

[3] W. Bender, D. Gruhl, and N. Morimoto, "Techniques for data hiding," in *Proc. SPIE*, vol. 2420, p. 40, Feb. 1995.

[4] J. Brassil, S. Low, N. Maxemchuk, and L. O'Gorman, "Electronic marking and identification techniques to discourage document copying," in *Proc. Infocom'94*, pp. 1278–1287.

[5] D. Boneh and J. Shaw, "Collusion-secure fingerprinting for digital data," in *Advances in Cryptology: Proc. CRYPTO'95*. New York: Springer-Verlag, 1995.

[6] G. Caronni, "Assuring ownership rights for digital images," in *Proc. Reliable IT Systems, VIS'95*. .

[7] I. J. Cox, S. Roy, and S. L. Hingorani, "Dynamic histogram warping of images pairs for constant image brightness," in *IEEE Int. Conf. Image Processing*, 1995.

[8] O. Faugeras, *Three Dimensional Computer Vision: A Geometric Viewpoint*. Cambridge, MA: MIT Press, 1993.

[9] A. Gersho and R. Gray, *Vector Quantization and Signal Compression*. Boston, MA: Kluwer, 1992.

[10] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. New York: Addison-Wesley, 1993.

[11] P. J. Huber, *Robust Statistics*. New York: Wiley,1981.

[12] N. Jayant, J. Johnston, and R. Safranek, "Signal compression based on models of human perception," in *Proc. IEEE*, vol. 81, no. 10, 1993.

[13] J. Kilian *et al.*, "Resistance of watermarked documents to collusional attacks," in preparation.

[14] E. Koch, J. Rindfrey, and J. Zhao, "Copyright protection for multimedia data," in *Proc. Int. Conf. Digital Media and Electronic Publishing*, 1994.

[15] E. Koch and Z. Zhao, "Toward robust and hidden image copyright labeling," in *Proc. 1995 IEEE Workshop on Nonlinear Signal and Image Processing*, June 1995.

[16] J. S Lim, *Two-Dimensional Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1990.

[17] F. T. Leighton and S. Micali, "Secret-key agreement without public-key cryptography," in *Proc. Cryptology*, 1993.

[18] B. M. Macq and J.-J. Quisquater, "Cryptology for digital TV broadcasting," in *Proc. IEEE*, vol. 83, pp. 944–957, 1995.

[19] K. Matsui and K. Tanaka, "Video-steganography," in *Proc. IMA Intellectual Property Project*, 1994, vol. 1, pp. 187–206.

[20] R. L. Pickholtz, D. L. Schilling, and L. B. Millstein, "Theory of spread spectrum communications—A tutorial," *IEEE Trans. Commun.*, vol. COMM-30, pp. 855–884, 1982.

[21] G. B. Rhoads, "Indentification/authentication coding method and apparatus," Rep. WIPO WO 95/14289, World Intellect. Property Org., 1995.

[22] W. F. Schreiber, A. E. Lippman, E. H. Adelson, and A. N. Netravali, "Receiver-compatible enhanced definition television system," U.S. Patent 5 010 405, 1991.

[23] C. W. Therrien, *Decision Estimation and Classification: An Introduction to Pattern Recognition and Related Topics*. New York: Wiley, 1989.

[24] K. Tanaka, Y. Nakamura, and K. Matsui, "Embedding secret information into a dithered multi-level image," in *Proc. 1990 IEEE Military Communications Conf.*, 1990, pp. 216–220.

[25] L. F. Turner, "Digital data security system," Patent IPN WO 89/08915, 1989.

[26] R. G. van Schyndel, A. Z. Tirkel, and C. F. Osborne, "A digital watermark," in *Int. Conf. Image Processing*, 1994, vol. 2, pp. 86–90.

Ingemar J. Cox (S'79–M'83–SM'95) received the Ph.D. degree from Oxford University, Oxford, U.K., in 1983.

From 1984 to 1989, he was a principal investigator in the Robotics Principles Department, AT&T Bell Laboratories, Murray Hill, NJ, where his research interests focused on issues of autonomous mobile robots. He joined NEC Research Institute, Princeton, NJ, as a senior research scientist in 1989. His principal research interests are broadly in computer vision, specifically tracking, stereo and 3-D estimation, and multimedia, especially image database retrieval and electronic watermarking for copyright protection.

Joe Kilian received the B.S. degree in computer science and in mathematics in 1985, and the Ph.D. in mathematics in 1989, both from the Massachusetts Institute of Technology, Cambridge.

He is a Research Scientist with NEC Research Institute, Princeton, NJ. His research interests are in complexity theory and cryptography.

F. Thomson Leighton received the B.S.E. degree in electrical engineering and computer science from Princeton University, Princeton, NJ, in 1978, and the Ph.D. degree in applied mathematics from the Massachusetts Institute of Technology (MIT), Cambridge, in 1981.

He is a Professor of applied mathematics and a member of the Laboratory for Computer Science (LCS) at MIT. He was a Bantrell Postdoctoral Research Fellow at LCS from 1981 to 1983, and he joined the MIT faculty as an Assistant Professor of applied mathematics in 1982. He is a leader in the development of networks and algorithms for message routing in parallel machines, particularly in the use of randomness in wiring to overcome problems associated with congestion, blocking, and faults in networks. He has published over 100 research papers on parallel and distributed computing and related areas. He is the author of two books, including a leading text on parallel algorithms and architectures.

Talal Shamoon (S'84–M'95) received the Ph.D. degree in electrical engineering from Cornell University, Ithaca, NY, in January 1995.

He joined the NEC Research Institute (NECI), Princeton, NJ, in December of 1994, where he held the title of Scientist. He joined the InterTrust STAR Laboratory, Sunnyvale, CA, in 1997, where he is currently a Member of the Research Staff working on problems related to trusted rights management of multimedia content. His research interests include algorithms for audio, image and video coding and processing, multimedia security, data compression, and acoustic transducer design. He has worked on high-fidelity audio coding and fast search algorithms for large image data bases. Since joining NECI, he has been actively involved in research on watermarking for multimedia systems.