CIRCUITS SYSTEMS SIGNAL PROCESSING
VOL. 17, NO 3, 1998, PP 361–389

A HOLOGRAPHIC TRANSFORM DOMAIN IMAGE WATERMARKING METHOD*

Alfred M. Bruckstein[1,2] *and Thomas J. Richardson*[1]

Abstract. A transform domain image tagging or *watermarking* method that survives image cropping (and, hence, is "holographic") was proposed at Bell Labs in September 1994. This report analyzes in detail the various properties of this method and introduces an optimal procedure for watermark recovery.

1. Introduction

This paper deals with a proposal for introducing imperceptible tags, called *watermarks*, into digitized images, enabling one to verify their status as authorized copies used by legitimate users. The tags are designed to carry information both on the source of the image (the copyright owner) and on the legitimate user who has purchased rights to the image, hence they enable tracking the illicit distribution of images. The tags or watermarks discussed are "holographic" in the sense that they can be detected even in small portions cropped from watermarked images. The method of tagging/watermarking discussed here was first proposed at Bell Labs in 1994 [2] but was never subjected to a thorough investigation, beyond some immediate feasibility tests. In this report we present results obtained with this watermarking method and propose an approach for the recovery of the watermark information from images compared to their original, untagged copies.

2. Holographic transform domain tagging

2.1. Watermarking method requirements

Let us first survey the requirements that must be met by a good solution of the

* Received June 30, 1997; revised August 18, 1997.
[1] Bell Laboratories, Murray Hill, New Jersey 07974.
[2] Computer Science Department, Technion, I.I.T, 32000, Haifa, Israel.

image watermarking problem. A solution to this problem will involve modifying the original (digitized) image $I(i, j)$ to produce another image $I^W(i, j)$ that will have embedded in it some information on the image source and its legal user. The amount of information in the watermark should be somewhere between 30 to 100 bits, and the requirements are as follows:

(1) The tags/watermarks should be imperceptible, i.e., I^W should appear very similar to the original image I.
(2) Tags should have the capability to carry sufficient information.
(3) Tags should not be easily identifiable from one or several differently tagged copies of the image.
(4) Tags should be easily recoverable from a tagged image and the original.
(5) Attempts to remove the tags or tamper with them should have noticeable ill effects on the image.
(6) Tags should not be wiped out by image modification/compression algorithms or by other casual image processing procedures.
(7) Tags should be distributed in the image plane and be recoverable from "arbitrary" portions of the image (should survive image cropping).

2.2. *Embedding watermarks in the frequency domain*

The proposal in [2] for watermarking is to introduce slight modifications of the image in some transform-domain image representation. For example, we can use the Fourier domain for tagging. An image $I(x, y)$ is transformed to $\tilde{I}(u, v)$ via the Fourier transform

$$\tilde{I}(u, v) = FT\{I(x, y)\} := \int\int I(x, y)e^{j2\pi(ux+vy)}\,dx\,dy,$$

where $\tilde{I}(u, v)$ is a complex bivariate function that can be represented as follows: $\tilde{I}(u, v) = M(u, v)e^{jP(u,v)}$ with $M(u, v) = |\tilde{I}(u, v)|$ the magnitude and $P(u, v) \in [0, 2\pi]$ the corresponding phase. Because $I(x, y)$ is real, we have $M(-u, -v) = M(u, v)$ and $P(u, v) = 2\pi - P(-u, -v)$.

It is well known that the "phase" image modifications are visually more perceptible than magnitude/amplitude modifications. Those latter modifications are well tolerated, leading to images that appear very similar to the originals. With this motivation, the following watermarking method was proposed in [2].

2.2.1. *Watermark method.*

(1) From $I(x, y)$ compute $\tilde{I}(u, v) = FT\{I(x, y)\}$.
(2) Embed a watermark (by any method) into the magnitude image $M(u, v)$, changing it to $M^W(u, v)$.
(3) Generate images to be distributed as

$$I^W(x, y) = FT^{-1}\{M^W(u, v)e^{jP(u,v)}\}.$$

The question remains: How should one modify $M(u, v)$ to get $M^W(u, v)$? One could say that we have simply not solved the problem as we are again facing an image watermarking problem, but the idea is this: we can use a method of modifying $M(u, v)$ that will meet much less stringent requirements than the original problem. So, we can, for example, modify $M(u, v)$ by multiplying it by a "watermark mask" $W_M(u, v)$ or by adding to it a "watermark mask" $W_A(u, v)$:

$$M^{W_M}(u, v) = W_M(u, v) \cdot M(u, v)$$
$$M^{W_A}(u, v) = W_A(u, v) + M(u, v).$$

In the first case $W_M(u, v)$ will have to be "close" to 1 everywhere, i.e., $W_M(u, v) = 1 + \varepsilon_M(u, v)$, and $\varepsilon_M(u, v)$ will be a function of the information bits $\{b_1, b_2, \ldots, b_N\}$ to be hidden into the watermark. In the second case $W_A(u, v)$ itself will have to be small for all (u, v) in order not to visibly perturb the watermarked images.

In the first case, that of multiplicative mask watermarking, we have that

$$I^W(x, y) = FT^{-1}\left\{M(u, v)e^{jP(u,v)} + \varepsilon_M(u, v)M(u, v)e^{jP(u,v)}\right\}$$
$$= I(x, y) + I(x, y) * FT^{-1}\{\varepsilon_M(u, v)\}.$$

In the second case we have

$$I^W(x, y) = FT^{-1}\left\{M(u, v)e^{jP(u,v)} + \varepsilon_A(u, v)e^{jP(u,v)}\right\}$$
$$= I(x, y) + FT^{-1}\{\varepsilon_A(u, v)e^{jP(u,v)}\}.$$

In both cases, in order to keep $I^W(x, y)$ real, we must choose to have $\varepsilon_{M/A}(u, v) = \varepsilon_{M/A}(-u, -v)$. For simplicity, we will discuss further only the first case.

We will consider piecewise constant functions $\varepsilon_M(u, v)$ that take values $0, \pm\varepsilon$. The information bits $b_1, \ldots, b_N$ will be encoded through the assignment of multiplicative mask values $1 \pm \varepsilon$ to various regions in the frequency domain. For example, in [2] it was proposed to have the bits $b_1, b_2, \ldots, b_N$ modulate a sequence of concentric rings in the frequency domain, as follows:

$$W_M(u, v) = 1 + \varepsilon(-1)^{b_i} \quad \text{for } \sqrt{u^2 + v^2} \in [r_i, r_{i+1}).$$

This method assigns the value $(1+\varepsilon)$ if $b_i = 0$ and the value $(1-\varepsilon)$ if $b_i = 1$ to the mask over a ring of spatial frequencies located between the radii r_i and r_{i+1}. Here the parameters $\varepsilon, r_1, r_2, r_3, \ldots, r_N$ are to be chosen so as to achieve imperceptibility (which requires small ε's) and good survival under various image modifications (which seems to require placing $r_1, \ldots, r_N$ into the lower frequencies). Clearly we could also have replaced the constant ϵ with a variable sequence of gains adapted to the frequency domain rings they modulate. Many other options for the design of mask functions are available: one could embed a variety of geometric designs or even a company logo into $W(u, v)$. Such designs could be of use when the purpose of the watermark is only to imperceptibly identify the source of the image. If we also need to have explicit bits encoded in $W(u, v)$ (in order, for example, to identify the recipient of the image), we can do this with gain sequences combined with a variety of geometric shapes.

2.3. *Watermark recovery*

Suppose we are given a watermarked version of the image $I^W(x, y)$ and we also have the original, master copy $I(x, y)$. Then, apparently, it is rather straightforward to recover the watermarks discussed in the previous section. In the multiplicative case we have, "ideally,"

$$\frac{\tilde{I}^W(u, v)}{\tilde{I}(u, v)} = \frac{M^W(u, v)}{M(u, v)} = W_M(u, v).$$

Therefore, it seems to be easy to recover the watermark, if one has access to the master copy $I(x, y)$ and an uncorrupted version of the watermarked image. However, there are several factors that make watermark recovery a nontrivial pursuit. One factor is the inherent image quantization. Images are digitized and stored as two-dimensional arrays of numbers, each represented with a finite number of bits. Therefore, even if watermarking is computed with high precision, the watermarked image will have to be requantized to the precision of the original, effectively injecting noise into the spectrum.

Other factors to be dealt with are various image-enhancement manipulations (contrast enhancement, sharpening/smoothing), image-editing operations (cropping, scale modifications, etc.), and the results of lossy image compression/decompression cycles that images can be subjected to by various users. The most serious threat to the embedded watermarks will come, of course, from a variety of attempts to deliberately remove or modify them. Watermarks, once embedded in an image, should be resilient against possible attacks by potential illicit users who may have something to gain from their removal or modification.

What makes the proposed watermarking method viable in view of these problems is the fact that the information bits are each redundantly encoded in regions of the transform/frequency domain. Any local modification in the transform domain is readily propagated and encoded over the entire image plane (hence the "holographic" property of the method), and each information bit is encoded over an extended region in the frequency domain, thereby achieving reliability through redundancy. Then the fact that the watermark geometry, i.e., the regions encoding the bits, can be made part of the encoded information makes it very difficult for an attacker to learn about the watermark bits.

We will consider using ratios of the form

$$\frac{\langle f(u, v), \tilde{I}^W(u, v)\rangle}{\langle f(u, v), \tilde{I}(u, v)\rangle}$$

as a means to watermark recovery, where, typically, we will have one or several distinct f functions for each bit supported in the area of the frequency domain reserved for that bit. Depending on how the watermark has been inserted into the image and on transformations to which $\tilde{I}^W$ may have been subjected, we will attempt to choose a collection of functions $\{f\}$ in an optimal way.

Note that the "ideal" recovery procedure is of the suggested form: we consider $f_{u_0,v_0}(u, v) = \delta(u - u_0, v - v_0)$ (for each point in the frequency domain (u_0, v_0)) and the preceding ratio becomes

$$\frac{\tilde{I}^W(u_0, v_0)}{\tilde{I}(u_0, v_0)} = 1 + \varepsilon(u_0, v_0),$$

and, under ideal conditions, we recover the watermark $\varepsilon(u, v)$ directly.

Before detailing effective ways to embed and recover watermarks, let us consider some of the image transformations that the watermarks will be expected to survive.

2.3.1. Effects of additive noise and linear operations on watermarks. Suppose we have generated a watermarked image $I^W(x, y)$ that is then subjected to a transformation as follows:

$$I_D^W(x, y) = A(x, y) * I^W(x, y) + B + N(x, y),$$

where B is a constant, $N(x, y)$ is a zero mean noise image, $A(x, y)$ is a smooth function, and $*$ denotes convolution. We then have

$$\begin{aligned}\tilde{I}_D^W(u, v) &= \tilde{A}(u, v) \cdot \tilde{I}^W(u, v) + B\delta(u, v) + \tilde{N}(u, v) \\ &= \tilde{A}(u, v)[1 + \varepsilon_M(u, v)]\tilde{I}(u, v) + B\delta(u, v) + \tilde{N}(u, v).\end{aligned}$$

The constant B renders the dc component effectively unrecoverable; hence, if we have reason to assume that $B \neq 0$ (and we usually have), then we should not hide any watermark information there! We may reasonably assume the function $\tilde{A}(u, v)$ to be smooth because nonsmooth effects may be absorbed into $\tilde{N}(u, v)$.

Many image transformations that are not necessarily linear filtering processes can, for our purposes, nevertheless be reasonably well modeled as such. Examples include printing, photocopying, and lossy compression.

2.3.2. Effects of cropping on watermarks. Suppose the original and watermarked image are supported on $[0, 1] \times [0, 1]$. Cropping the image corresponds to multiplying $I(x, y)$ by $\text{Rect}(c_x(x - x_0),\ c_y(y - y_0))$, where $\text{Rect}(x, y) = 1_{[0,1]\times[0,1]}$. We have

$$FT(\text{Rect}(x, y)) = e^{-\imath\pi(u+v)}\text{sinc}\,\pi u/\text{sinc}\,\pi v,$$

hence

$$\begin{aligned}&FT\left(I(x, y)\,\text{Rect}(c_x(x - x_0), c_y(y - y_0))\right) \\ &\quad = \frac{1}{c_x c_y}\tilde{I}(u, v) \otimes e^{-i2\pi(c_x x_0 + c_y y_0)} e^{-i\pi(\frac{u}{c_x} + \frac{v}{c_y})}\text{sinc}\,\frac{\pi u}{c_x}\,\text{sinc}\,\frac{\pi v}{c_y}.\end{aligned}$$

Thus, the effect of cropping on the Fourier transform is to convolve it with a complex smoothing function. To the extent that the constant regions of our watermark are large compared to the main peak of the sinc function, the watermark will

survive, i.e., we will have

$$(1+\varepsilon(u,v))\tilde{I}(u,v) \otimes \frac{e^{-i2\pi(c_x x_0 + c_y y_0)}}{c_x c_y} e^{-i\pi(\frac{u}{c_x}+\frac{v}{c_y})} \operatorname{sinc}\frac{\pi u}{c_x} \operatorname{sinc}\frac{\pi v}{c_y}$$

$$\approx (1+\varepsilon(u,v))\left[\tilde{I}(u,v) \otimes \frac{e^{-i2\pi(c_x x_0 + c_y y_0)}}{c_x c_y} e^{-i\pi(\frac{u}{c_x}+\frac{v}{c_y})} \operatorname{sinc}\frac{\pi u}{c_x} \operatorname{sinc}\frac{\pi v}{c_y}\right]$$

for those u, v where $\varepsilon(u, v)$ is locally constant on a scale comparable to $c_x \times c_y$. This approximate equality will not hold near the boundaries where the function $\varepsilon(u, v)$ changes discontinuously. Thus we see that the watermark is largely expected to survive cropping if the regions over which $\varepsilon(u, v)$ is constant are sufficiently large. Of course, when we attempt to recover the watermark, we must compare against an identical cropping of the original image. We will not discuss here the rather minor effects of replacing the Fourier transform by the discrete Fourier transform.

2.3.3. The effects of lossy compression. Lossy compression of $I(x, y)$ involves replacing $I(x, y)$ with a version $I^{cp}(x, y)$ that requires fewer bits to encode than $I(x, y)$, but is nonetheless similar to it in some subjective/objective distance measure. It is difficult to evaluate the influence of various compression algorithms; however, we can state, in general, that it will involve filtering out visually "imperceptible" frequency components of $I(x, y)$. When embedding a watermark into the image these general facts must be given careful consideration. We can model the compression effects as a combination of linear filtering and additive noise: a model we have already discussed. In fact, when we compare a JPEG-compressed image to its original version, we realize that the model of a multiplicative mask in the frequency domain is quite reasonable (see Figure 7 for experimental results).

2.3.4. Attacks on watermarks. When proposing a watermarking procedure we must keep in mind that, given some reasonable economical/financial motivations, there will be serious and professionally well-informed attempts to tamper and modify and/or remove the tags embedded in images. We may assume that there will be legal arrangements in place, requiring each copyrighted document to have a legible watermark embedded in it, hence it will be the job of the illegal users to modify, not remove, the existing watermarks. If one wanted to disturb the watermark embedded by the method proposed herein, one could simply multiply the FT of the image by a random pattern $W_{\text{rand}}(u, v)$. This would generate an image with a watermark of $W_M(u, v) \cdot W_{\text{rand}}(u, v)$ from which it would not be possible, in general, to recover $W_M(u, v)$. However, this would yield an image that could not be legally used because it would lack a valid watermark. In order to generate through such an attack from $W_M(u, v)$ a watermark that would be legitimate, albeit different, one would have to know the geometry of the watermarks and multiply the transformed image by a $W_{\text{attack}}(u, v)$ that was adapted to the frequency domain geometry of the watermarks.

In the literature, people also mention "collusion attacks," in which several watermarked images are used to learn about the watermarks and subsequently modify

them. Indeed, if one has two watermarked images one could, assuming identical geometry, approximately recover, say, various ratios $\{\frac{W_{M_i}(u,v)}{W_{M_j}(u,v)}\}$ that would lead to knowledge about the watermark geometry. Therefore, part of the security in the watermarking process proposed must also come from freedom to *parameterize* the geometry of the spectral masks employed.

3. Optimal watermark recovery/detection

In this section we address the question of optimal watermark recovery from watermarked images, possibly degraded and filtered, based on their comparison to a "golden" or "reference" original.

Within the context of watermarking in the Fourier transform domain, there are essentially two strategies that can be used to enable the watermark to survive various expected image transformations. These are

A. Embedding a watermark in such a way that its recovery is not affected by the transformations,
B. Attempting to identify or model the transformation so as to compensate for it prior to watermark recovery.

In the case of cropping, for example, approach B seems to be required in any case. One must precisely locate the cropped portion in the original image. This may be a nontrivial task if the image has also been resampled.

With regard to linear filtering, we have primarily considered approach A in this paper. Our intention, in general, is to guard against a spatially smooth scaling of the Fourier transform and additive noise.

A potentially effective B-type strategy is the following. We may leave various regions of the frequency domain unaltered, i.e., we set $\varepsilon(u, v) = 0$ there. Given a modified watermarked image, we may "sample" the spectral modification in the unmarked regions and appropriately interpolate to obtain estimates of the modification in the marked regions. These estimates may be used to approximately invert or model the modifications in the watermarked regions. It may then be possible to use much less geometric redundancy in the watermark and much more error-correcting coding.

3.1. General considerations

The key to our optimal watermark recovery procedure involves the following general considerations from estimation/detection theory. Suppose we have a set of complex numbers $\{s_i\}_{i=1,2,\ldots,k}$ and a set of observations q_i given by

$$q_i = \alpha s_i + n_i,$$

where α is a real value, and n_t are independent realizations of a complex noise random variable.

In this setting we can address the following questions:

(1) What is the optimal estimator of α given $\{q_i\}_{i=1,2,\ldots,k}$, or
(2) If $\alpha = 1 + \varepsilon$ or $\alpha = 1 - \varepsilon$ what is the optimal decision on whether α is higher or lower than one?

We shall assume that n_i are independent complex Gaussian random variables with mean 0 and variance $2\sigma^2$. To answer our first question, we can denote by $p(\mathbb{Q} \mid \alpha)$ the likelihood of seeing the data given some value of α and maximize this with respect to α to obtain the maximum likelihood (ML) estimate of α. We have

$$p(\mathbb{Q} \mid \alpha) = \prod_{i=1}^{K} \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2\sigma^2}|q_i - \alpha s_i|^2},$$

and here $p(\mathbb{Q} \mid \alpha)$ is maximized if

$$P(\alpha) := \sum_{i=1}^{K} |q_i - \alpha s_i|^2$$

is minimized. Hence the optimal estimator for α is

$$\hat{\alpha}_{\text{opt(ML)}} = \frac{\sum_i \text{Re}\{q_i^* s_i\}}{\sum_i s_i^* s_i},$$

where z^* denotes the complex conjugate of z. For the second question, we deal with a hypothesis-testing problem. If we assume that the $\alpha_+ = 1+\varepsilon$ and the $\alpha_- = 1-\varepsilon$ cases have equal prior probabilities, then the optimal (hypothesis-testing) decision process proceeds via the following likelihood ratio test:

$$\lambda(\mathbb{Q}) = \frac{P(\mathbb{Q} \mid \alpha_+)}{P(\mathbb{Q} \mid \alpha_-)} = \frac{P(\mathbb{Q} \mid 1+\varepsilon)}{P(\mathbb{Q} \mid 1-\varepsilon)}.$$

This yields

$$\frac{\prod_{i=1}^{k} e^{-\frac{1}{2\sigma^2}|q_i - \alpha_+ s_i|^2}}{\prod_{i=1}^{k} e^{-\frac{1}{2\sigma^2}|q_i - \alpha_- s_i|^2}} \underset{\alpha_-}{\overset{\alpha_+}{\gtrless}} 1,$$

which, by taking logs, is seen to be equivalent to

$$\sum |q_i - \alpha_+ s_i|^2 \underset{\alpha_+}{\overset{\alpha_-}{\lessgtr}} \sum |q_i - \alpha_- s_i|^2,$$

and this reduces to

$$\frac{\sum_i \text{Re}(q_i^* s_i)}{\sum_i |s_i|^2} \underset{\alpha_-}{\overset{\alpha_+}{\gtrless}} \frac{\alpha_+ + \alpha_-}{2}.$$

If $\alpha_+ = 1 + \varepsilon$ and $\alpha_- = 1 - \varepsilon$, then the threshold is 1. We obtain the result that the optimal decision rule proceeds via

$$\hat{\alpha}_{\text{opt(ML)}} \underset{1-\varepsilon}{\overset{1+\varepsilon}{\gtrless}} 1,$$

not an unexpected result. Hence we can state the following result: the optimal way to recover α from the $\{q_l\}$ measurements is by calculating $\hat{\alpha} = \frac{\text{Re}(\sum_l q_l^* s_l)}{\sum_l s_l^* s_l}$ and, if a priori we know that α takes on some known values α_+, α_-, we have to compare $\hat{\alpha}$ to the average value of α_+ and α_-.

We shall use this general result to optimally detect/recover the watermark embedded in the image. For example, assuming that the embedded bits have been selected from some code, we may use the values of α_+ and α_- in a soft decoding procedure.

3.2. The watermark recovery procedure

We shall assume, as before, that the process of embedding a watermark proceeds via multiplication of $\tilde{I}(u, v)$ by a mask $W_M(u, v) = 1 + \varepsilon_M(u, v)$. The watermarked image $I^W(x, y)$ is first quantized to fit the way images are represented in the computer. Then the quantized image may undergo cropping, compression/decompression (via JPEG, say), some smoothing and dynamic range corrections, and corruption by some additive noise. The resulting quantized/corrupted $I^W(x, y)$ then becomes $\tilde{I}_D^W(u, v)$ in the frequency domain, and we shall regard each frequency component of $\tilde{I}_D^W(u, v)$ as a complex observation of the corresponding spectral component of $\tilde{I}(u, v)$.

To apply the analysis considered in Section 3.1 to the general linear filtering model discussed in Section 3.2.1, it is necessary that the multiplicative scaling of the spectrum $\tilde{A}(u, v)$ be (approximately) constant and *known*. This may not be possible in general.

As discussed before, assuming that the unknown multiplicative scaling mask $\tilde{A}(u, v)$ is smooth, it would be feasible to estimate it by reserving a small, approximately uniformly distributed, portion of the frequency domain for estimation purposes. The spectrum would be unaltered there, i.e., we would have $\tilde{I}^W(u, v) = \tilde{I}(u, v)$, so that an estimate of $\tilde{A}(u, v)$ could be formed and extrapolated over the entire frequency domain. We have not experimented with this approach, although we consider it viable. Rather, we will consider methods for modulating the watermark data so as to ameliorate the effects of smooth but unknown linear filtering (corresponding to multiplicative spectral scaling by $\tilde{A}(u, v)$).

If, for example, we know a priori that over the regions allocated to each bit in the watermark the unknown filtering factor changes by only a very small amount, i.e., we have $\tilde{A}(u, v) \simeq \beta = \text{const}$ there, then we could encode the bit of information over that particular region by using the following idea: partition the region R_n

allocated to bit b_n into two disjoint subregions $R_{n,1}$ and $R_{n,2}$ and encode b_n as follows:

$$\text{if } b_n = 0 \quad \begin{cases} W(u,v) = 1+\varepsilon & (u,v) \in R_{n,1} \\ W(u,v) = 1-\varepsilon & (u,v) \in R_{n,2} \end{cases}$$

$$\text{if } b_n = 1 \quad \begin{cases} W(u,v) = 1-\varepsilon & (u,v) \in R_{n,1} \\ W(u,v) = 1+\varepsilon & (u,v) \in R_{n,2} \end{cases}.$$

In this case, we may recover the bit b_n from the "spectral" observations provided by the degraded $\tilde{I}^W(u,v)$ over R_n in a straightforward way; we'll simply have to decide whether

$$\beta\hat{\alpha}_{(R_{n,1})} \underset{-}{\overset{+}{\gtrless}} \beta\hat{\alpha}_{(R_{n,2})},$$

using the ideas presented in Section 3.1. In general, it will not be optimal to decode each bit separately when the embedded bits are chosen from a code.

Let q_i denote the values of $\tilde{I}_D^W(u,v)$ and let s_i denote the values of $\tilde{I}(u,v)$, where i indexes (u,v). Let N_j denote the indices associated to $R_{n,j}$, $j = 1,2$, respectively. Let us assume that $q_i = \beta\alpha s_i + n_i$, where the n_i are i.i.d. complex Gaussians and $\alpha \in \{1-\varepsilon, 1+\varepsilon\}$. Calculating likelihood ratios, we find that the optimal decision rule when β is known is given by

$$\sum_{i\in N_1} \mathrm{Re}(q_i^* s_i) + \beta \sum_{i\in N_1} |s_i|^2 \underset{b_n=0}{\overset{b_n=1}{\gtrless}} \sum_{i\in N_2} \mathrm{Re}(q_i^* s_i) + \beta \sum_{i\in N_2} |s_i|^2. \tag{3.1}$$

Note that in the case where $\sum_{i\in N_1} |s_i|^2 = \sum_{i\in N_2} |s_i|^2$, this reduces to

$$\frac{\sum_{i\in N_1} \mathrm{Re}(q_i^* s_i)}{\sum_{i\in N_1} |s_i|^2} \underset{b_n=0}{\overset{b_n=1}{\gtrless}} \frac{\sum_{i\in N_2} \mathrm{Re}(q_i^* s_i)}{\sum_{i\in N_2} |s_i|^2}. \tag{3.2}$$

We see that in this case it is optimal to form estimates, as described in Section 3.1, one each for $R_{n,1}$ and $R_{n,2}$, and then to take the difference. If we choose $R_{n,1}$ and $R_{n,2}$ contiguous and of equal area, then we can typically expect that we will have $\sum_{i\in N_1} |s_i|^2 \simeq \sum_{i\in N_2} |s_i|^2$ as needed to get (3.2).

We want the watermark to survive cropping, so we cannot afford to make the areas $R_{n,1}$ and $R_{n,2}$ too small. Thus, $\tilde{A}(u,v)$ may not be sufficiently close to a constant. Fortunately, the "differential encoding" idea that we have presented can be generalized.

Consider, for example, a smooth function f supported on a rectangular domain R. Suppose that we subdivide R into m uniform columns and n uniform rows, where nm is even, creating a grid, and color the grid in a black-and-white checkerboard pattern. Let f_w denote the average of f over the white squares, and let f_b denote the average of f over the black squares. As m and n tend to infinity, f_w and f_b both tend to f_R, the average of f over R. In particular, observe that $(1+\varepsilon)f_w - (1-\varepsilon)f_b$ tends to $2\varepsilon f_R$. More specifically, if we expand f in a Taylor

series about the center of R and choose $m = n = 2$, then $f_w - f_b$ depends only on terms of order greater than or equal to 2.

Consider the decision rule given by (3.2) and let us write $q_i = \beta_i s_i$, where β_i discretizes (in the plane) $\tilde{A}(u, v)$ and assume that N_1 and N_2 are given by a (2×2) checkerboard as just described. Assume that $|s_i|^2$ is a discretized smooth function S in the plane and consider the center of the checkerboard to be the origin. It follows that the sums $\sum_{i \in N_1} |s_i|^2$, $\sum_{i \in N_2} |s_i|^2$ are both equal to the 0th-order term of S plus other terms depending only on terms of order at least 2. If we assume that $|s_i|^2$ can be modeled as S plus zero-mean i.i.d. noise, then these sums have an additional random component whose variance is inversely proportional to $|N_1|$ and $|N_2|$, respectively.

If we assume that $\sum_{i \in N_1} |s_i|^2 = \sum_{i \in N_2} |s_i|^2$, then we see that the decision rule (3.2) amounts to comparing two sums. If we assume that $A(u, v)S$ is smooth, then we see that the sums are equal up to second order terms. Even without assuming that $\sum_{i \in N_1} |s_i|^2 \neq \sum_{i \in N_2} |s_i|^2$, it follows that the difference of the sums depends only on terms of order at least 2.

Other variations on the decision rule are possible. For example, we may estimate β as

$$\frac{1}{2}\left(\frac{\sum_{i \in N_1} \mathrm{Re}(q_i^* s_i)}{\sum_{i \in N_1} |s_i|^2} + \frac{\sum_{i \in N_2} \mathrm{Re}(q_i^* s_i)}{\sum_{i \in N_2} |s_i|^2}\right)$$

and then substitute the estimate into (3.1). In our experiments the resulting rule gave the same bit errors as those obtained using (3.2) and reported in the next section.

4. Implementation and tests

The watermarking and watermark recovery ideas that we have discussed were extensively tested on a set of three (512×512) grey-level (8-bit) images, chosen from an art library database. The images are cropped grey-level versions of paintings by Renoir, Michelangelo, and Botticelli (see Figure 1). The method chosen for watermark embedding was the multiplicative one discussed in Section 2. The magnitude of the (512×512) Fourier transform of these images was multiplied by a mask function $M(u, v \mid B)$ dependent on a vector B of 120 binary digits $\{b_{m,n} \in \{+1, -1\}, m = 1, 2, \ldots, 8, n = 1, 2, \ldots, 15\}$. The geometry of the mask was chosen to be a very simple one: if one regards the frequency domain corresponding to the (512×512) images as $[-1, 1] \times [-1, 1]$, and introduces polar coordinates (r, θ), then $\mathcal{R}_{n,m}$, the region reserved for embedding bit $b_{m,n}$, is defined by

$$\mathcal{R}_{m,n} = \left\{(r, \theta) : r \in [(m-1)\frac{1}{15}, m\frac{1}{15}),\ \theta \in [(n-1)\frac{\pi}{8}, n\frac{\pi}{8})\right\}.$$

(Recall that the $M(u, v \mid B)$ must satisfy the symmetry property $M(u, v \mid B) = M(-u, -v \mid B)$ so that each region $\mathcal{R}_{m,n}$ is also duplicated by reflection about the origin.) Each region $\mathcal{R}_{m,n}$, rectangular in polar coordinates, was subdivided into a (2×2) rectangular array for differential encoding, as described in Section 3.2. See Figure 1 for examples of $M(u, v \mid B)$.

The watermarked version of each test image $I(x, y)$ was defined to be (requantized, by rounding and clipping)

$$I^W(x, y) = FT^{-1}\{(1 + \varepsilon M(u, v \mid B))I(u, v)\}.$$

In Figures 2, 3, and 4 we show, respectively, the Botticelli, the Michelangelo, and the Renoir test images watermarked with an arbitrary B vector and various gain factors ($\varepsilon = 0.0, 0.05, 0.1, 0.2$). We can see that for $\varepsilon = 0.05$ (the value chosen for subsequent tests) and $\varepsilon = 0.1$ the watermarked images are virtually indistinguishable from their originals. Figure 1 shows watermarked versions of each of the three test images with pseudo-random watermark information bit sequences B and $\varepsilon = 0.5$. The ideal local watermark recovery is not perfect because of the requantization of the watermarked images. However, we see that 120 bits of data are perfectly recovered by the optimal watermark estimation procedure in spite of this.

After these initial tests addressing the effects of requantization alone on the watermark recovery, several others were performed in order to see how the 120 bits of watermark data are affected by various degradations of the watermarked images. The degradations and modifications of the images under which we need to exhibit sufficiently precise survival of the watermark are

(1) cropping portions of the image
(2) lossy compression-decompression cycles (with the standard JPEG algorithm)
(3) changes in contrast
(4) printing on standard laser printers and rescanning the image (with possible cropping).

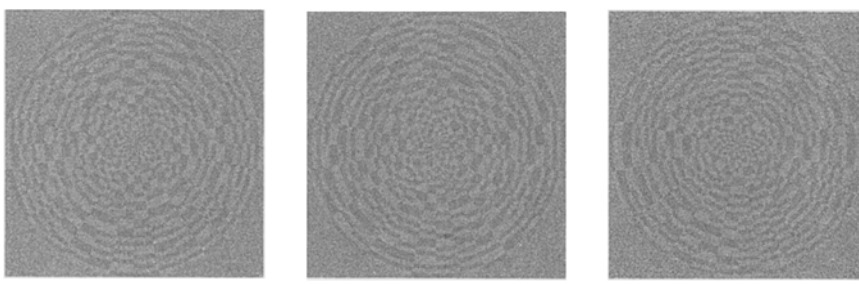
In order to test for watermark survival after printing-cropping-rescanning degradation it is necessary to solve the rather delicate problem of registering the rescanned and possibly cropped image with the original. Recall that the watermark detection mechanism assumes the availability of two images (or image portions) in perfect registration.

Let f represent an image that may be a degraded subimage of a watermarked image whose original we denote by g. Let Ω denote the rectangle over which f is defined. We register f by finding (using a multiresolution hill climbing algorithm) an affine transformation Ψ that minimizes the following functional:

$$\int_\Omega \frac{g(\Psi(x)) - \bar{g}}{\|g(\Psi(x)) - \bar{g}\|} \frac{f(x) - \bar{f}}{\|f(x) - \bar{f}\|} dx,$$

Original images: Botticelli, Michelangelo, Renoir.

Watermarked images



Watermark recovery, pointwise ratios

Figure 1. Original and watermarked ($\varepsilon = 0.05$) images.

Original Image: Botticelli

Watermarked with ε=0.05

Watermarked with ε=0.1

Watermarked with ε=0.2

Figure 2. Results for Botticelli image: watermarking at various intensities.

Original Image: Michelangelo

Watermarked with ε=0.05

Watermarked with ε=0.1

Watermarked with ε=0.2

Figure 3. Results for Michelangelo image: watermarking at various intensities.

Original Image: Renoir

Watermarked with ε=0.05

Watermarked with ε=0.1

Watermarked with ε=0.2

Figure 4. Results for Renoir image: watermarking at various intensities.

where

$$\bar{g} := \int_{\Omega} g(\Psi(x))\,dx\,,$$

$$\bar{f} := \int_{\Omega} f(x)\,dx\,.$$

Once the desired Ψ is found, we resample the image appropriately.

In actual implementations the 120 bits would be chosen from a code. Perhaps 30 bits would actually identify the user. In this situation it is preferable to use soft decoding.

4.1. *Watermark recovery from cropped portions of watermarked images*

We ran several cropping tests on the three images shown in Figure 1, with cropped portions of size 256×256. Typically, we could recover the bits with 2/3 errors only. Figure 5 shows the watermark recovery results. (Here we assumed perfect alignment/registration with the original.) The few errors that do occur lie in the low-frequency components. This is undoubtedly due to the relatively small area reserved for bits embedded into these components.

4.2. *Watermark recovery from images that underwent JPEG compression*

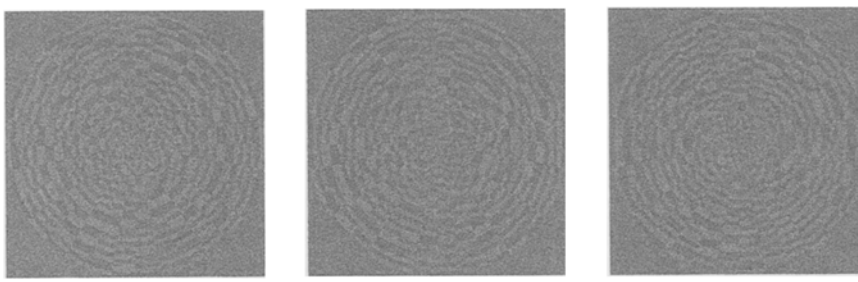
Here we took the watermarked images ($\varepsilon = 0.05$) and compressed them using a standard JPEG algorithm with a quality factor of 20% producing considerable visible degradations and blocking effects in the images, as seen in Figures 6 and 7.

In all three images we could recover correctly the first (i.e., corresponding to low frequency) 55 bits of the watermark string B, and at least 90 of the 120 bits we correctly recovered. With higher quality factors the results would be better. As seen in the spectral image of watermark recovery, the high frequencies are quite strongly affected by the compression—not surprisingly. Hence the bits embedded in this range of the spectrum should not be expected to survive. What is important to notice, however, is that because of the "differential" encoding of each bit, the low-range encoded information remains intact in spite of the fact that the compression process effectively scales the spectrum with a varying gain envelope (see Figure 7).

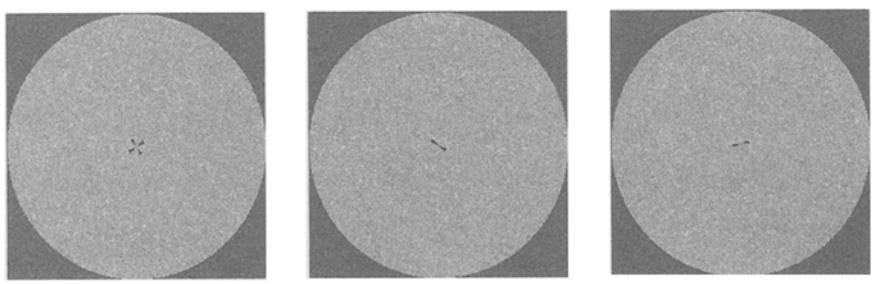
4.3. *Watermark recovery under printing and cropping followed by image rescanning*

This cycle of degradations is the most severe that our watermarks are required to survive. First we printed the watermarked originals on a 600-dpi laser printer, then

Cropped portions of watermarked image: Botticelli, Michelangelo, Renoir.



Watermark recovery: average values (note reduced contrast due to blurring)



Watermark recovery: error locations; totals: 2,1,1 resp.

Watermark recovery: error locations; totals: 2, 1, 1 respectively

Figure 5. Results for cropped images ($\varepsilon = 0.05$).

Undegraded Watermarked Image: Michelangelo

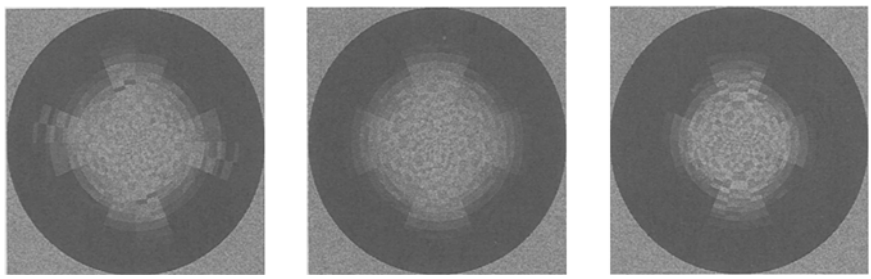
JPEG 20% compressed Michelangelo

JPEG 20% compressed Renoir

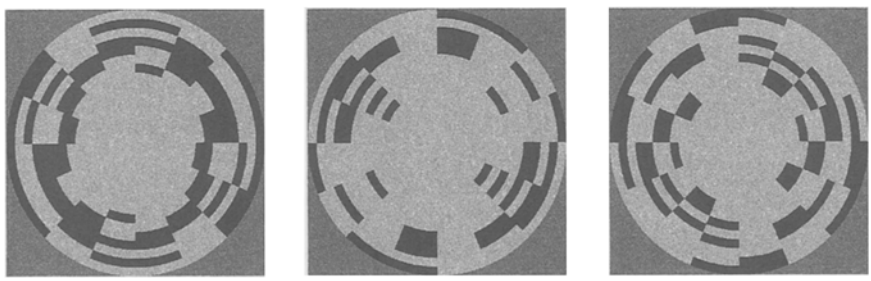
JPEG 20% compressed Botticelli

Figure 6. Watermarked images after 20% quality JPEG compression ($\varepsilon = 0.05$).

JPEG compressed (20%) watermarked image: Botticelli, Michelangelo, Renoir.



Watermark recovery: average values (note high frequency loss)



Watermark recovery, error locations; totals, 29,18,24 resp.

Figure 7. Results for images after 20% quality JPEG compression ($\varepsilon = 0.05$).

we scanned the images with a typical color image scanner at a resolution of 120 dpi. This was done for both $\varepsilon = 0.05$ and $\varepsilon = 0.1$. The scanning for $\varepsilon = 0.1$ had a different brightness setting than that for $\varepsilon = 0.05$. We then appealed to our general-purpose multiresolution image registration procedure (see Section 4) to generate an optimally registered (512×512) image for each of the three rescanned images. The results for $\varepsilon = 0.05$ are shown in Figure 8, and those for $\varepsilon = 0.1$ are shown in Figure 10.

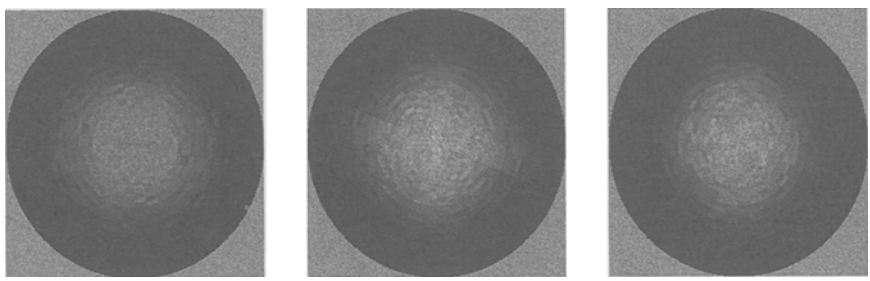
We also cropped portions slightly larger than 1/4 of the image from the scanned images and optimally registered these portions. We then resampled according to the original image and then cropped a (256×256) portion from the registered sections and a corresponding portion from the originals. On these images we then ran our optimal watermark recovery algorithm. The results for $\varepsilon = 0.05$ are shown in Figures 8 and 9, and the results for $\varepsilon = 0.1$ are shown in Figures 10 and 11. As we see, in this case, with $\varepsilon = 0.05$, we can typically recover 105 bits correctly (out of the 120 that were embedded into the watermarked images) for uncropped images and 95 bits when using the optimally registered, redigitized cropped image sections of 1/4 area. With $\varepsilon = 0.1$ we did remarkably better, having only 0, 0, 2 errors for the three uncropped images and 0, 2, 11 errors from the three cropped images. Some of this improvement can be attributed to the larger value of ε; however, we believe that some of the improvement resulted from the different brightness setting of the scanner which better reproduced the contrast in the scanned image. This is indicated also by the fact that it is the Renoir image that produced the larger number of errors in the $\varepsilon = 0.1$ images. The Renoir image is significantly brighter than the others, so it may have suffered more dynamic range compression under the brighter scanning. Thus, significant improvement may be possible with higher gray-level resolution scanners.

From these experimental results we may conclude that even in the worst case under such severe degradations, we can expect to recover 90 bits out of the 120, mostly the ones embedded into the lower-frequency range.

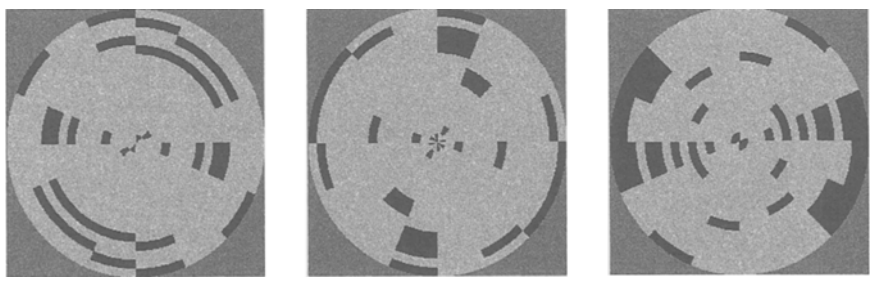
4.4. Watermark recovery from images that underwent changes of contrast

Here we adjusted the contrast of the image in a nonlinear way by adjusting the intensity values using a gamma value of 2. The results are shown in Figure 12. We observe that the pointwise ratio exhibits apparently random noise. The overall brightening of the image reduces contrast and hence the energy in much of the spectrum. Because of the "differential" encoding, all of the bits in each image were nevertheless recovered correctly.

Cropped portions of printed/scanned watermarked image: Botticelli, Michelangelo, Renoir



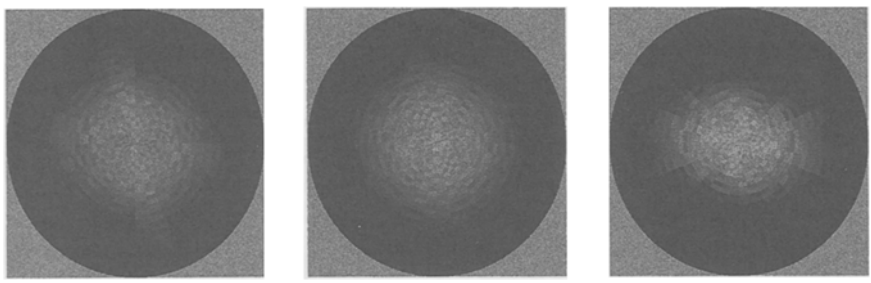
Watermark recovery: average values (note high frequency loss)



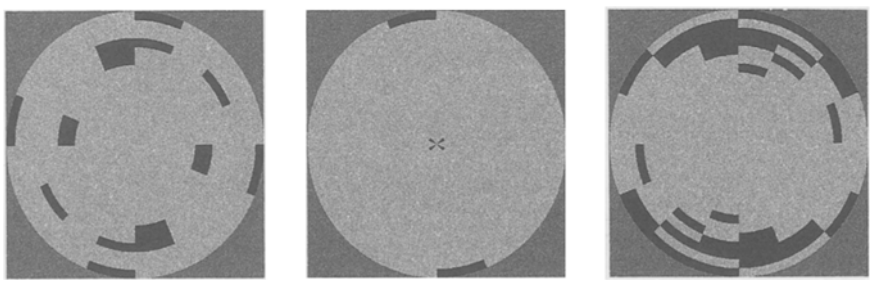
Watermark recovery: error locations: totals: 16,16,24 resp.

Figure 8. Results for printed/scanned and cropped images with $\varepsilon = 0.05$.

Printed/scanned watermarked image: Botticelli, Michelangelo. Renoir.



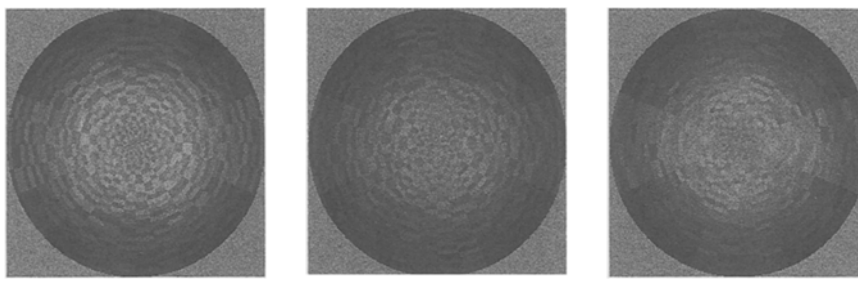
Watermark recovery: average values (note high frequency loss)



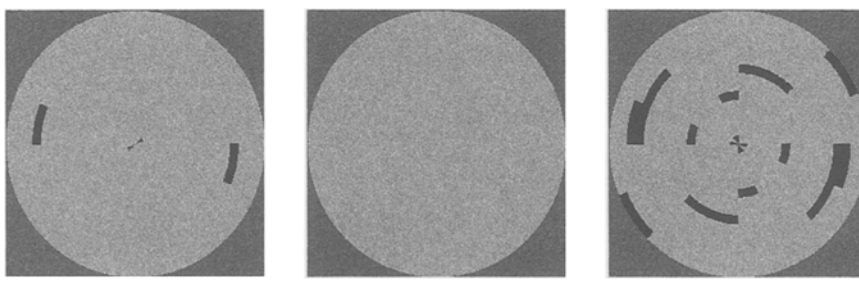
Watermark recovery. error locations; totals: 9,3,17 resp.

Figure 9. Results for printed/scanned images with $\varepsilon = 0.05$.

Cropped portions of printed/scanned watermarked image: Botticelli, Michelangelo, Renoir.



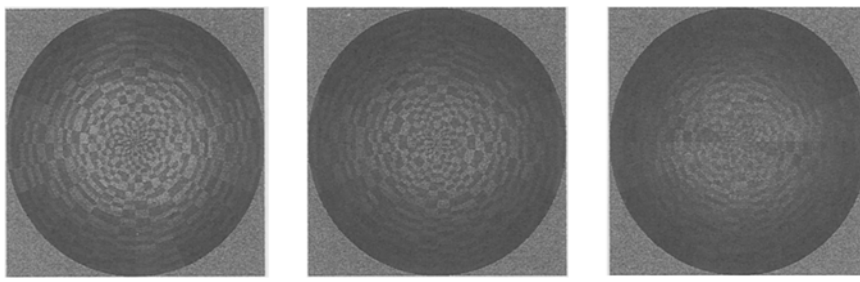
Watermark recovery: average values (note high frequency loss)



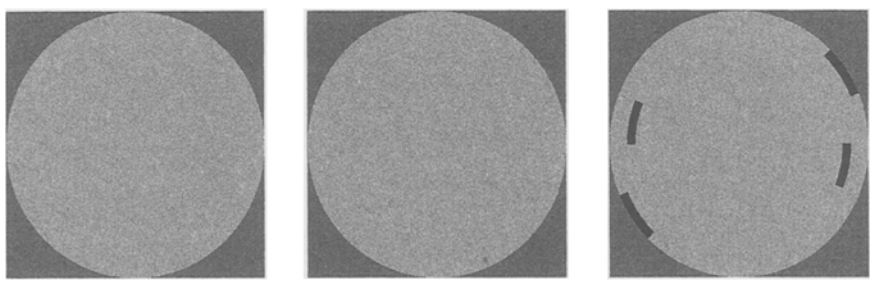
Watermark recovery: error locations; totals. 2,0,11 resp.

Figure 10. Results for printed/scanned and cropped images with $\varepsilon = 0.1$.

Printed/scanned watermarked image: Botticelli, Michelangelo, Renoir.



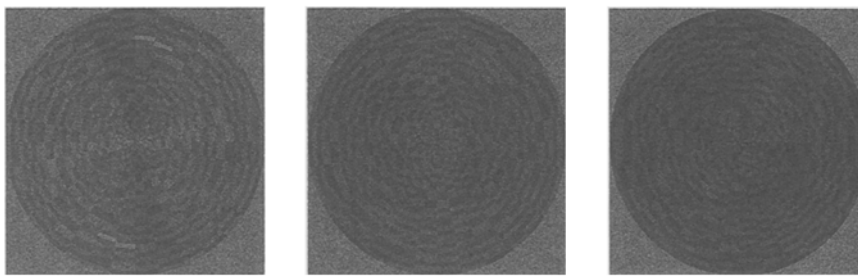
Watermark recovery: average values (note high frequency loss)



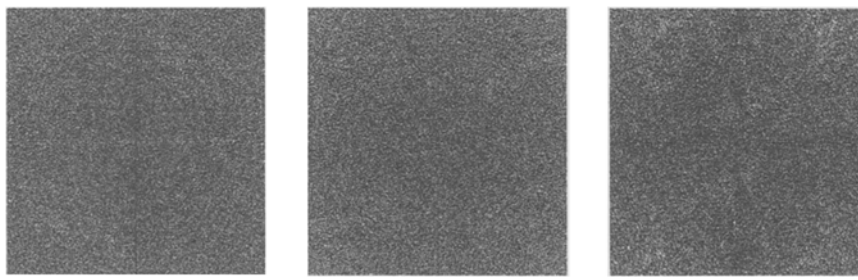
Watermark recovery: error locations, totals. 0,0,2 resp.

Figure 11. Results for printed/scanned images with $\varepsilon = 0.1$.

Gamma curve 2 intensity adjustment· Botticelli, Michelangelo. Renoir.



Watermark recovery: average values (note darkness due to reduced contrast)



Watermark recovery:pointwise values

Figure 12. Results for gamma adjusted images (all zero erros) with $\varepsilon = 0.05$.

5. Conclusions and practical recommendations from the watermarking experiments

The preceding, rather extensive, but clearly not exhaustive, set of experiments lead us to the following conclusions and recommendations:

(1) *For watermarking use the lower part of the spectrum, without the vicinity of the dc component.*

In our experiments we chose, rather arbitrarily, to encode the information in 120 sections in the entire spectral domain, with approximately equal energy, under the assumption that the spectrum is decaying as $1/r$ in amplitude from the center. This rather arbitrary choice, can of course be customized and optimized, and we could have decided to use a lower section of the spectral domain, selected fewer bits to encode, and made sure that all regions chosen had exactly the same energy. For practical implementations of the watermarking method we propose to indeed invest some effort in such customization and optimization processes.

(2) *Use differential encoding of the bits as proposed.*

As the experiments have shown, we can indeed model various nonlinear degradations, by a rather smooth multiplicative mask in the frequency domain. The differential encoding idea seems to handle such degradations very well. This idea, too, can be refined and adapted to more specific information, if it becomes available, on the various degradations following compression/decompression cycles, printing and rescanning, etc.

(3) *Use error correction for the bit strings embedded as watermarks.*

As we have seen in the rough experiments that we carried out with the (arbitrary) choice of 120 bits to be embedded in the entire frequency domain corresponding to the (512×512) images, we can typically recover these bits with less than 25 errors. However, most of the errors occur in the high-frequency range. It may be desirable, therefore, not to use these frequencies. However, even in the case of a relatively crude choice of geometry and bit density and spectral domain watermarking, an error-correcting code could safely enable us to embed about 40 error-free information bits into the images on which we experimented. This is more than enough for practical purposes.

There arises in this context the following issue: How can the recovered watermark bits be used to guarantee the identity associated with the image? If we can recover 40 bits error free with high probability, then, because 20 bits should be sufficient to encode the identity, we can guarantee that all valid 40-bit watermarks are well separated in Hamming distance. Thus if the recovered 40 bits are valid, then the probability that those are the correct 40 bits, uniquely identifying the source and recipient of the image, will be very high. This probability value attached to the recovered watermark could be further refined by using soft decoding, for example.

6. Discussion

At the time the memo "On tagging images" [2] was written and circulated (September 1994), there were very few works published on this topic. The works of Caronni [3] and O'Gorman [5] proposed to embed watermarks in the spatial domain by slightly enhancing or depressing the image grey levels. The works of Matsui and Tanaka [7] and of Dautzenberg et al. [1] discussed the possibilities of embedding watermarks in the frequency domain, but only in the context of JPEG-style block coding. In [2] it was stressed that one can rely on the transform domain to achieve "holographic watermarking" with the property that each little piece of the image will carry a (perhaps) somewhat degraded, but recognizable, version of the "tag" or watermark.

Since September 1994 the field of image watermarking has rapidly developed, and today there are entire sessions devoted to it at various image processing conferences. The most remarkable recent contribution to this area was due to Cox et al. [4] in 1995. In work first reported as an NEC Research Report and already published in several other places, these authors also realized that the frequency domain is the natural place to do watermarking. They use the DCT for that purpose, and propose to do watermarking by randomly choosing a set of unique frequency components into which to encode a pseudo-random (noise-like) watermark vector, to be then recovered by correlation. The frequency components chosen are perceptually significant (which means they reside in the low-frequency range) to ensure that the relevant information will not be lost due to cropping, compression, etc. The work of Cox and his colleagues, although similar in spirit to ours, differs from it significantly. In [2] and in this paper we have proposed to directly embed bits of watermark information in conjunction with a carefully chosen geometrically defined and parameterized pattern on the frequency plane. Each of these bits is encoded redundantly in a region of the transform plane and then recovered using an optimal detection procedure. It is our pleasure, however, to say that [4], [2], and the present paper have all clearly established the idea that the frequency domain representation coefficients of the image (low-frequency range) are the natural candidates to carry the imperceptible watermark information. We think (and we are clearly biased in our views) that our method, first proposed in 1994 and thoroughly tested herein in association with the optimal watermark recovery procedure described, has several appealing features from a practical point of view:

(1) The geometry of the watermark can be used to carry information (e.g., a company logo).

(2) The redundant robust and direct coding of up to 40 bits of information is possible through the use of error-correcting codes.

(3) The method provides for easy encoding/decoding.

(4) It is difficult to detect/replace the watermark without explicit knowledge of the watermark geometry.

We refer the reader to the work of Cox et al. [4] and to [6] for a clear and up-to-date survey of watermarking efforts since 1994, and to [2] for several ideas for extensions and further developments like pyramidal (wavelet-based) watermarking and using the visual system properties in the design of the watermarks. Clearly, we could and should vary the amount by which the frequency domain regions are emphasized/deemphasized according to some perceptibility metric on the frequency components of images. In fact, several efforts in this direction have already been made by several researchers, see, e.g., [8].

In summary, after extensive testing, we see our watermarking method, in conjunction with the optimal watermark recovery method, as a highly competitive proposal to be considered by the industry in its attempts to settle upon a standardized solution to the digital image authentication problem.

Acknowledgments

A. Bruckstein thanks Larry O'Gorman and Nick Maxemchuck for bringing this topic to his attention during his visit at Bell Laboratories in the summer of 1994. This was the motivation for the work reported in [2].

Many interesting and illuminating discussions on these issues with A. N. Netravali are also acknowledged.

References

[1] F. M. Boland, J. J. K. O'Ruanaidh, and C. Dautzenberg, Watermarking digital images for copyright protection, *Proceedings of the 5th International Conference on Image Processing and its Application*, Edinburgh, Scotland, IEE Publication no. 410, pp. 326–330, 4–6 July 1995.

[2] A. M. Bruckstein, On tagging images, AT&T Bell Labs, Murray Hill, NJ, *Internal Memorandum*, September 16, 1994.

[3] G. Caronni, Assuring ownership rights for digital images, *Asia Script*, 94.

[4] I. J. Cox, J. Kilian, T. Leighton, and T. Shamoon, Secure spread-spectrum watermarking for Multimedia, *NEC Research Report*, 1995.

[5] L. O'Gorman, Watermarks for security and authentication of digital pictures, Patent application, AT&T Bell Labs, Murray Hill, NJ, June 20, 1994/95.

[6] Information hiding, *Proceedings of an Isaac Newton Institute Workshop*, Univ. of Cambridge, Cambridge, England, May 1996.

[7] K. Matsui and K. Tanaka, Video-stegonography: How to secretly embed a signature in a picture, *IMA Intellectual Property Project Proceedings* 11, 1994.

[8] M. D. Swanson, B. Zhu, and A. H. Tewfik, Transparent robust image watermarking, *Proc. of ICIP96*, Lausanne, Switzerland, 1996.