Keywords: Copyright Protection, Digital Watermark, Spread Spectrum,

A Secure Robust Digital Image Watermark

Joseph J. K. Ó Ruanaidh and Shelby Pereira

Computer Vision Group
Centre Universitaire d'Informatique
CUI Université de Genève
24 rue Général Dufour
CH 1211 Genève 4
Switzerland

ABSTRACT

Digital watermarks have been proposed as a method for discouraging illicit copying and distribution of copyright material. This paper presents a new approach for the secure and robust copyright protection of digital images. The digital watermarks described in this paper are designed to be, as far as possible, invariant against image transformations such as rotation, translation, scaling and cropping. We concentrate especially on the desirable properties of the Fourier Transform and propose a novel technique based on *an invisible template* which allows us to reverse many of the effects of image processing on the digital watermark. Robustness of the watermark to operations such as lossy compression is achieved by using a perceptually adaptive spread spectrum communications approach, in which a spread spectrum signal is embedded in selected components of the magnitude spectrum of the image. The keys used to embed the spread spectrum signal are generated, certified, authenticated and securely distributed using a public key infrastructure containing an electronic copyright office and a certification authority. The security architecture used for this purpose is also outlined.

1. INTRODUCTION

Digital media have become common and have increasingly taken over and have extended the applications of traditional analog media. There are a great number of technical reasons for favoring digital media. Infrastructure such as computers, printers and high rate digital transmission facilities are becoming very inexpensive, widely available and more widespread. Digital networks also provide an efficient cost-effective means of distributing digital media. The popularity of the World Wide Web has clearly demonstrated the commercial potential of the digital multimedia market and consumers are investing heavily in digital audio, image and video recorders and players. Unfortunately however, digital networks and multimedia also afford virtually unprecendented opportunities to pirate copyrighted material. Digital storage and transmission make it trivial to quickly and inexpensively construct *exact* copies. The idea of using a robust digital watermark to detect and trace copyright violations has therefore stimulated significant interest among artists and publishers. As a result, digital image watermarking has recently become a very active area of research. Techniques for hiding watermarks have grown steadily more sophisticated and increasingly robust to lossy image compression and standard image processing operations, as well as to cryptographic attack.

Many of the current techniques for embedding marks in digital images have been inspired by methods of image coding and compression. Information has been embedded using the Discrete Cosine Transform (DCT)[1,2] Discrete Fourier Transform magnitude and phase,[3] Wavelets,[1] Linear Predictive Coding[4] and Fractals.[5] The key to making watermarks robust has been the recognition that in order for a watermark to be robust it must be embedded in the *perceptually significant* components of the image.[1,2] The term "perceptually significant" is somewhat subjective but it suggests that a good watermark is one which takes account of the behaviour of human visual system. Objective criteria for measuring the degree to which an image component is significant in watermarking have gradually evolved from being based purely on energy content[1,2] to statistical[6] and psychovisual[7,8] criteria.

Digital watermarking is fundamentally a problem in digital communications.[1,9,2] In parallel with the increasing sophistication in modelling and exploiting the properties of the human visual system, there has been a corresponding development in communication techniques. Early methods of encoding watermarks were primitive and consisted of no more than incrementing an image component to encode a binary '1' and decrementing to encode a '0'.[10,1] Tirkel and Osborne[11] were the first to note the applicability of spread spectrum techniques to digital image watermarking. Since then there has been an increasing use of spread spectrum communications in digital watermarking. It has several advantageous features such as cryptographic security,[11,12,2] and is capable of achieving error free transmission of the watermark near or at the limits set by Shannon's noisy channel coding theorem.[1,9]

Other author information: (Send correspondence to Shelby Pereira)
Shelby Pereira: E-mail: pereira@cui.unige.ch

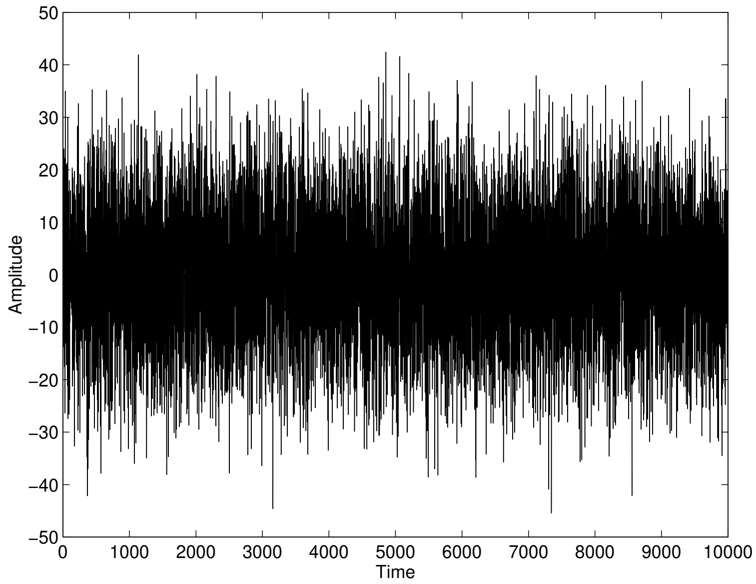

Figure 1. An example of a spread spectrum signal used as a digital watermark.

Spread spectrum is an example of a symmetric key[13] cryptosystem. System security is based on proprietary knowledge of the keys (or the seeds for pseudorandom generators) which are required to embed, extract or remove an image watermark.

The ability of humans to perceive the salient features of an image regardless of changes in the environment is something which humans take for granted.[14,15] We can recognize objects and patterns independently of changes in image contrast, shifts in the object or changes in orientation and scale. Gibson[16] makes the hypothesis that the human visual system is strongly tied to the ability to recognize invariants. It seems clear that an embedded watermark should have the same invariance properties as the image it is intended to protect. In this respect, we propose that an image watermark should be, so far as possible, encoded to be *invariant* to image transformations. We shall also demonstrate how image invariants might be used to construct watermarks that are unaltered by some of the most basic operations encountered in image processing; namely rotation, translation and changes of scale.

2. SPREAD SPECTRUM

Pickholtz et al.[17] define spread spectrum communications as follows:

> Spread spectrum is a means of transmission in which the signal occupies a bandwidth in excess of the minimum necessary to send the information; the band spread is accomplished by a code which is independent of the data, and a synchronized reception with the code at the receiver is used for despreading and subsequent data recovery.

Spread spectrum systems are capable of approaching the Shannon limit for reliable communication. The fundamental information theoretic limits to reliable communication and its implications to digital watermarking have been discussed by some authors.[1,9] Note that the smaller is the number of bits of core information or "payload" contained in a watermark, the greater is the chance of it being communicated without error.

Cox et al[2] recover a watermark by subtracting an original image from the marked image and explicitly computing the correlation between the (noise corrupted) watermark recovered from the image with a set of perfect watermarks stored in a database. This is a very robust technique for watermark recovery but it is not very useful in practice because of the need for access to both the original image and a database of perfect watermarks and the large amount of computation required. In this paper, the approach is similar to other spread spectrum approaches in that the watermark is embedded in the form of a pseudorandom sequence. However the approach is different to that of Cox et al in that it does not require access to a database of watermarks and is not particularly expensive computationally. In common with other spread spectrum techniques, in order to embed a mark, or to extract it, it is important to have access to the key which is simply the seed used to generate pseudorandom sequences. In the case of a public watermarking scheme the key is generally available and may even be contained in publically available software. In a private watermarking scheme the key is proprietary. From the point of view of embedding watermarks in documents given the keys or seeds the sequences themselves can be generated with ease. A mark may be embedded or extracted by the key owner which, in our model, is the Copyright Holder. In this form spread spectrum is a symmetric key cryptosystem. The infrastructure required to generate, issue and store the keys is outlined in section 3 of this paper.

One proviso in the use of a spread spectrum system is that it is important that the watermarking process incorporate some non-invertible step which may depend on a private key or a hash function of the original image. Only in this way can true ownership of the copyright material be resolved. Otherwise, one can compute a "counterfeit original" – in fact, it is even possible for an attacker to make it appear that the true original contains her watermark.[18] The obvious solution is to employ *oblivious* watermarks (meaning that an original image is not required to extract the embedded message) but there is the disadvantage that oblivious watermarking is not as robust. Recent work by Cox[19] has pointed out other flaws in Digital watermarking which make the need for a secure Digital Watermark Infrastructure such as that mentioned in section 3 all the greater.

2.1. CDMA coding of digital watermarks

A method for encoding binary messages which can later be recovered given knowledge of the key used is described here. Suppose we are given a message which, without loss of generality, is in binary form $b_1, b_2 \cdots b_L$ where each b_i is a bit. This can be written in the form of a sequence of symbols $s_1, s_2 \cdots s_M$, most generally by a change in a number base from 2 to B with $L \leq M \log_2 B$. The conversion from base 2 to a base which is a power of two is trivial. The next stage is to encode each symbol s_i in the form of a zero mean pseudorandom vector of length N. To encode the first symbol a pseudorandom sequence $\vec{v}$ of length $N + B - 1$ is generated. To encode a symbol of values where $0 \leq s < B$ the elements $v_s, v_{s+1} \cdots v_{s+N}$ are extracted as a vector $\vec{r_1}$ of length N. For the next symbol another independent zero mean pseudorandom sequence is generated and the symbol encoded as a random vector $\vec{r_2}$. Each successive symbol is encoded in the same way. Note that even if the same symbol occurs in different positions in the symbol sequence no collision is possible because the random sequences used to encode them are different — in fact they are statistically independent. Finally the entire sequence of symbols is encoded as the summation:

$$\vec{m} = \sum_{j=1}^{L} \vec{r_j} \tag{1}$$

The pseudorandom vector $\vec{m}$ is decoded by generating all of the random vectors $\vec{r_j}$ in turn and recovering the symbols which give the largest value of cross correlation. One form of pseudorandom sequence that is used is an m-sequence* but this is not material to the issue since any "good" generator will do. A good spread spectrum sequence is one which combines desirable statistical properties such as uniformly low cross correlation with cryptographic security. In other words, the specific choice of method for generating the pseudorandom sequence has direct implications to the reliability and security of the embedded mark. Pseudorandom number generators described in watermarking literature include Gold Codes, Kasami codes, m-sequences, Legendre sequences and perfect maps.[20,11,21,12] In addition, one may use two dimensional or higher dimensional arrays[12] in place of the one dimensional pseudorandom vectors described in the communications system above.

One interesting point is that for M sufficiently large the statistical distribution of the message $\vec{m}$ should approach a Gaussian distribution. This follows from the Central Limit Theorem. A Gaussian distributed watermark has the advantage that it is more difficult to detect. The variance increases with order M — in other words, the expected peak excursion of the sequence is only order $M^{\frac{1}{2}}$.

Figure 1 shows a spread spectrum signal $\vec{s}$ composed of a linear combination of L random vectors $\vec{r_j}$ as given by equation 1. Each random vector is specifically chosen to represent a particular symbol occupying a certain position in the message. A symbol may be composed of any number of bits. In our case each symbol is eight bits long and the number of random vectors L is nineteen. This is a form of Direct Sequence Code Division Multiple Access spread spectrum communications. The encoded message in Figure 1 reads "This is a watermark".

This form of spread spectrum is resistant to cropping (providing it is resynchronised), non-linear distortions of amplitude and additive noise. Also, if it has good statistical properties it should be mistaken for noise and go undetected by an eavesdropper.

There are however some drawbacks to using direct sequence spread spectrum. Although a spread spectrum signal as described above is extremely resistant to non-linear distortion of its amplitude and additive noise it is also intolerant of timing errors (i.e. getting the starting point for decoding wrong). Synchronization is of the utmost importance during watermark extraction. If watermark extraction is carried out in the presence of the original image then synchronization is relatively trivial. The problem of synchronizing the watermark signal is much more difficult to solve in the case where there is no original image. If the watermarked image is translated, rotated and scaled then synchronization necessitates a search over a four dimensional parameter space (X-offset, Y-offset, angle of rotation

*One effective approach is to use a zero mean m-sequence of length $N = 2^p - 1$ where p is the order of the primitive polynomial used to generate the m-sequence. Each symbol can be represented as one of the $B = N$ cyclic shifts of this sequence.

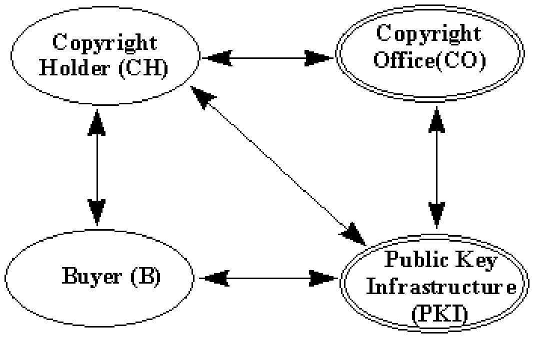

Figure 2. *Communication channels between identified parties.*

and scaling factor). The search space grows even larger if one takes into account the possibility of shear and a change of aspect ratio. In this paper, the aim is to investigate the possibility of using invariant representations of a digital watermark to help avoid the need to search for synchronization during the watermark extraction process.

3. THE COPYRIGHT NETWORK

We envision the watermark system operating an open environment like the Internet with different interconnected computers. Users can be located anywhere and can sell or buy images. In order to receive legally binding watermarks, the Copyright Holder (H) sends copyright information and the image information to the Copyright Certificate Center (C). After having received a copyright certificate from C, the copyright holder can sell his digital images, for example, via a secure image shopping mall, to an image buyer (B). The Public Key Infrastructure (PKI) supports the distribution of authentic public keys between all parties which are needed for mutual authentication and secure communication. The communication channels between the parties are shown in figure 2.

3.1. Copyright Protection and Image Owner Authentication

Depending on the proof-level to be provided for copyright protection, our approach provides three increasing levels of reliability, namely: individual copyright protection, copyright protection with registered cryptographic keys and copyright protection with C on the basis of registered cryptographic keys. The present method detects the Image Authentication Data (IAD) as the payload of a watermark. In order to embed or extract a watermark, it is necessary to know the exact values of the seed used to produce pseudo random sequences used to encode a watermark. The seeds are considered to be cryptographic keys for watermark generation and verification. System security is therefore based on proprietary knowledge of private keys, which provide in addition the necessary security parameters needed for a secure communication (mutual authentication, integrity, confidentiality, non-repudiation) in the trading process of digital images. Because spread spectrum signals are statistically independent (and therefore virtually orthogonal), the present method and apparatus encodes more than one watermark in an image at the same time, namely private, detection and public watermarks. The *detection watermark* is embedded under a fixed random seed which allows H to efficiently search for his images on the Internet.

The *public watermark* indicates that the image is copyright material and provide information on true ownership. At the same time there is a secure *private watermark* whose secrecy depends on the private key of the H. Since the public key of the H is registered, H can prove that he is the only person in the possession of the adequate private key and is therefore the generator of the private watermark. The system also provides the secure registration (mutual authentication, integrity, non-repudiation) of watermark encoded images (data sets) at C. The stego image is registered at C and a signed digital copyright certificate is generated by C. If an unauthorized third party has also encoded watermarks in the same image, conflicting claims in copyright disputes can be resolved, as only one of the two parties has a copyright certificate for the image containing only its watermark. The other party, who redistributed the original watermarked image, has only a certificate on the image where both watermarks are embedded and thus can be identified as the cheating party.

Watermark protection with registered cryptographic keys and C based copyright protection are based on a PKI. The PKI issues on request public key certificates such as X.509 certificates, containing the public key of the party, its distinguished name and a time stamp. Every certificate is signed with the PKI's private key and trust is built on the validity of the authentic copy of the PKI's public key (we assume that the public key of the PKI is accessible, authentically distributed and verifiable by every party). The precise details of the PKI and the remaining members of the Copyright Network are described by Petersen et al..[22]

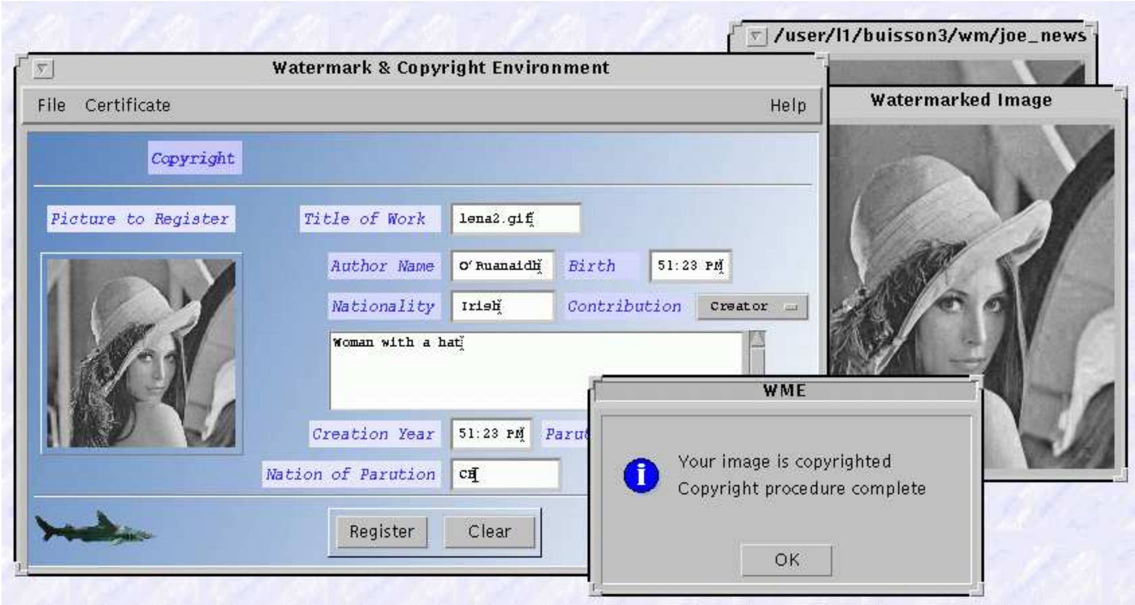

Figure 3. *The watermarked image of Lena being registered for copyright protection using a Java console*

To demonstrate the feasibility of the approach, a Java/C++ based copyright protection and authentication environment for digital images has been implemented. An example of this copyright protection environment in action is shown in Figure 3. The PKI, the H, C and the IB application processes all implement a Graphical User Interface and a server, supporting both console users and other requests through a socket interface.

Note that this security architecture for the copyright protection of digital images can also be extended to other data such as video, audio and binary data. We are currently investigating new spread spectrum techniques for the watermarking of video data.

4. THE FOURIER TRANSFORM

In this section, we shall describe the Fourier transform in some detail highlighting those properties which make it particularly suitable to digital image watermarking.

4.1. Definition

Let the image be a real valued continuous function $f(x_1, x_2)$ defined on an integer-valued Cartesian grid $0 \leq x_1 < N_1, 0 \leq x_2 < N_2$.

The Discrete Fourier Transform (DFT) is defined as follows:

$$F(k_1, k_2) = \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} f(x_1, x_2) e^{-j2\pi x_1 k_1/N_1 - j2\pi x_2 k_2/N_2} \tag{2}$$

The inverse transform is

$$f(x_1, x_2) = \frac{1}{N_1 N_2} \sum_{k_1=0}^{N_1-1} \sum_{k_2=0}^{N_2-1} F(k_1, k_2) e^{j2\pi k_1 x_1/N_1 + j2\pi k_2 x_2/N_2} \tag{3}$$

The DFT of a real image is generally complex valued. This leads to magnitude and phase representation for the image:

$$A(k_1, k_2) = [F(k_1, k_2)] \tag{4}$$

$$\Phi(k_1, k_2) = \angle F(k_1, k_2) \tag{5}$$

4.2. General Properties of the Fourier Transform

It is extremely instructive to study the effect of an arbitary linear transform on the spectrum of an image. From this study we will conclude that it is possible to undo the effect of any linear transformation on an image, even if the image is cropped.

Once $N_1 = N_2$ (i.e. square blocks) the kernel of the DFT contains a term of the form:

$$x_1 k_1 + x_2 k_2 = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} \tag{6}$$

If we compute a linear transform on the spatial coordinates:

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \rightarrow \mathbf{T} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{7}$$

then one can see that the value of the DFT will not change† if:

$$\begin{bmatrix} k_1 \\ k_2 \end{bmatrix} \rightarrow \left(\mathbf{T}^{-1}\right)^T \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} \tag{8}$$

4.3. FFT: Rotation

Consider a rotation matrix

$$\mathbf{T} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{9}$$

Therefore,

$$\left(\mathbf{T}^{-1}\right)^T = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{10}$$

Rotating the image through an angle θ in the spatial domain causes the Fourier representation to be rotated through the same angle.

$$\begin{aligned} &F(k_1\cos\theta - k_2\sin\theta, k_1\sin\theta + k_2\cos\theta) \\ &\quad \leftrightarrow f(x_1\cos\theta - x_2\sin\theta, x_1\sin\theta + x_2\cos\theta) \end{aligned} \tag{11}$$

Note that the grid is rotated so the value of the image at the new grid points may not be defined. The value of the image at the nearest valid grid point can be estimated by interpolation.

4.4. FFT: Scale

Consider a scaling matrix

$$\mathbf{T} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \tag{12}$$

Therefore,

$$\left(\mathbf{T}^{-1}\right)^T = \begin{bmatrix} \frac{1}{\lambda_1} & 0 \\ 0 & \frac{1}{\lambda_2} \end{bmatrix} \tag{13}$$

Hence, scaling the axes in the spatial domain causes an inverse scaling in the frequency domain.

$$\frac{1}{\rho}F\left(\frac{k_1}{\rho}, \frac{k_2}{\rho}\right) \leftrightarrow f(\rho x_1, \rho x_2) \tag{14}$$

4.5. FFT: Skewing

Consider a skewing matrix

$$\mathbf{T} = \begin{bmatrix} 1 & 0 \\ \alpha & 1 \end{bmatrix} \tag{15}$$

Therefore,

$$\left(\mathbf{T}^{-1}\right)^T = \begin{bmatrix} 1 & -\alpha \\ 0 & 1 \end{bmatrix} \tag{16}$$

Therefore, skewing the y coordinates of the spatial coordinates inverse skews the x coordinates of the frequency coordinates (and vice versa).

†The DFT will be invariant except for a scaling factor which depends on the Jacobian of Transformation, namely the determinant of the transformation matrix T.

4.6. FFT: Translation

Shifts in the spatial domain cause a linear shift in the phase component.

$$F(k_1, k_2) \exp\left[-j(ak_1 + bk_2)\right] \leftrightarrow f(x_1 + a, x_2 + b) \tag{17}$$

Note that both $F(k_1, k_2)$ and its dual $f(x_1, x_2)$ are periodic functions so it is implicitly assumed that translations cause the image to be "wrapped around". We shall refer to this as a *circular translation* or a cyclic shift. From property 17 of the Fourier transform it is clear that spatial shifts affect only the phase representation of an image. This leads to the well known result that the magnitude of the Fourier transform is a circular translation invariant.

It is less well known that it is possible to derive invariants based on the phase representation. To do this involves eliminating the translation dependent linear term from the phase representation. Brandt and Lin[23] present two such translation invariants, namely the *Taylor invariant* which removes the linear phase term in the Taylor expansion of the phase and the *Hessian invariant* which removes this linear phase term by double differentiation.

We shall see in section 4.7 that properties 14 and 11 allow one to extend the basic translation invariants to cover changes of rotation and scale.

4.7. The Fourier-Mellin Transform

The basic translation invariants described in section 4.6 may be converted to rotation and scale invariants by means of a *log-polar map.*

Consider a point $(x, y) \in \Re^2$ and define:

$$\begin{aligned} x &= e^{\mu} \cos\theta \\ y &= e^{\mu} \sin\theta \end{aligned} \tag{18}$$

where $\mu \in \Re$ and $0 \leq \theta < 2\pi$. One can readily see that for every point (x, y) there is a point (μ, θ) that uniquely corresponds to it.

The new coordinate system has the following properties:

Scaling is converted to a translation.

$$(\rho x, \rho y) \leftrightarrow (\mu + \log\rho, \theta) \tag{19}$$

Rotation is converted to a translation.

$$\begin{aligned} &(x\cos(\delta) - y\sin(\delta), x\sin(\delta) + y\cos(\delta)) \\ &\quad \leftrightarrow (\mu, \theta + \delta) \end{aligned} \tag{20}$$

At this stage one can implement a rotation and scale invariant by applying a translation invariant in the log-polar coordinate system. Taking the Fourier transform of a log-polar map is equivalent to computing the Fourier-Mellin transform:

$$F_M(k_1, k_2) = \int_{-\infty}^{\infty} \int_0^{2\pi} f(e^{\mu}\cos\theta, e^{\mu}\sin\theta) \exp\left[i(k_1\mu + k_2\theta)\right] d\mu\, d\theta \tag{21}$$

The modulus of the Fourier-Mellin transform is rotation and scale invariant.

4.8. What does a log polar map look like?

Figure 4 shows the effect of a log polar map on the standard image Lena. The region in the centre of the image is emphasised (i.e. oversampled) at the expense of regions further away from the centre. This point is amply illustrated by considering the inverse log polar map in Figure 5 computed from Figure 4.

4.9. Rotation, Scale and Translation Invariance

Ó Ruanaidh and Pun[24] discuss the possibility of embedding a watermark in a domain that is invariant to rotation, scale and translation transformations using a combination of Fourier Transforms and a Log Polar map. A prototype watermarking system illustrated in Figure 6 was proposed.

To give a concrete example of its application, consider a copy of a watermarked image placed on a scanner from which we wish to extract an embedded mark. The image may be reduced or increased in size and will be, more often than not, at an angle of $\pm\epsilon$, $\pm 90 \pm \epsilon$ or even $180 \pm \epsilon$ degrees where $\pm\epsilon$ is some small random angle. The image is also likely to be translated. Using the invariants derived above it should be possible to to extract an embedded mark regardless of orientation, scale or position.

Figure 4. *Log polar map of Lena.*

Figure 5. *Inverse Log Polar map of Lena.*

5. ROBUST OBLIVIOUS CROPPING INVARIANT IMAGE WATERMARKING ALGORITHM

In this section, we will describe in detail the approach used to embed a watermark that is simultaneously resistant to lossy image compression such as JPEG, invariant to cropping and which is *oblivious* (i.e. there is no need for an original image (*cover image*) to extract the digital watermark from the watermarked image (*stego image*).

5.1. Oblivious Watermarking

It is well known that transform based image compression techniques favour low frequencies, in the sense that low frequency content in the original image is better preserved in a compressed image. For this reason, it would seem that a low frequency watermark would be better. However, in oblivious watermarking it is necessary to avoid low frequencies for embedding information because the image itself interferes with the watermark at those frequencies.[9] Fortunately, a compromise is possible: a judicious selection of a band of frequencies leads to a watermark that is both oblivious and is sufficiently resistant to lossy image compression. One helpful factor is that there are relatively

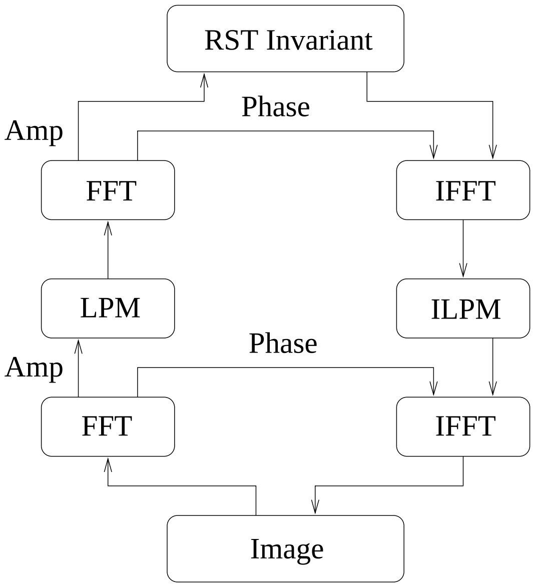

Figure 6. *A diagram of a prototype RST invariant watermarking scheme. FFT and IFFT denote a Fast Fourier Transform and its inverse and LPM and ILPM denote a Log Polar Map and its inverse respectively.*

few low frequency components in which to embed a spread spectrum signal. Using midband frequencies actually improves the robustness of the mark because of the increased redundancy of the encoding of the payload.

5.2. Cropping Invariance and Phase

One feature of translation invariants developed using the magnitude of the Fourier transform is that they are invariant to circular translations (or cyclic shifts). This is used to construct watermarks that are invariant to cropping. Given that the watermark information is only embedded in the magnitude spectrum the question arises as to the rôle of the phase spectrum. Lim[25] discusses the relative importance of magnitude and phase and concludes that, of the two, the phase spectrum is more important to the integrity of the image.

Phase therefore plays a very important rôle in the watermarking embedding algorithm. There are two cases:

1. The trivial case arises if the image is marked as a single block. In this case the phase is left unchnaged.

2. If the image is very large, it may not be feasible to compute the Fourier Transform. In this case, a copy of the watermark is embedded in each block. Suppose that the watermark in a standard size block will be of the form: $T = [AB; CD]$ where the submatrices A, B, C and D are of arbitrary size. A circular translation of such a watermark is of the form: $S = [DC; BA]$. The original stego image is tiled with watermarks in the pattern $[TTTT; TTTT; TTTT]$. A little thought demonstrates that a cropped section of the matrix will carry a watermark in the form $[SSSS; SSSS; SSSS]$. When reading the watermark of the cropped image, each block carries the watermark S. Since S is a circular transform of T, the magnitude spectrum will be the same for both T and S. Note, however, that the cover image is not tiled, only the watermark is. Therefore, while cropping merely induces a circular translation of the watermark in each block, the change of image in each block is not a circular translation. However, if one simply uses the local phase in each block then cropping invariance is lost. The reason for this is *phase cancellation* due to the effect of mixing adjacent blocks. This problem is readily resolved by enforcing that the phase spectrum be the same in each block. We denote this *the reference phase.* The reference phase should be random to avoid creating visible patterns in the image. A compromise between the conflicting goals of preserving the image spectrum and the need to use a reference phase to preserve cropping invariance is possible. The watermark uses local image phase but the sign of the watermark is modulated according to the difference in image phase and reference phase. If the two phases are opposing (i.e. separated by more than π radians) the watermark is inverted, otherwise it is left alone. One constraint on this solution is that the magnitude spectrum is always positive which means that it is not always possible to change the sign of the watermark.

5.3. Embedding a Watermark

Given a cover image the steps for embedding a watermark in the image are as follows:

1. If the image is a color image, then compute the luminance component (for example, by simply replacing each pixel by $g/2 + r/3 + b/6$, where g, r and b are its green, red and blue components) and use these values for the following calculations.

2. Map the image luminance levels (or gray levels for a black and white image) to a perceptually "flat" domain by replacing them with their intensity. The logarithm is an intuitively good choice because it corresponds to the Weber-Fechner law which describes the response of the human visual system to changes of luminance. This step ensures that the intensity of the watermark is diminished in the darker regions of the image where it would otherwise be visible.

3. Compute the FFT (Fast Fourier Transform) of each block. From the real and imaginary components obtained in this way, calculate the corresponding magnitude and phase components. The magnitude components are translation invariant and will therefore be used in the following modulation steps. (However, it is possible to derive translation invariants from the phase spectrum as well, which could also be modulated).

4. Select the magnitude components to be modulated. To encode a message m of length N, a total number of N components are modulated. In non-oblivious watermarking, any components can be modulated. For oblivious watermarking, because of the interference of the cover image with the watermark, the largest magnitude components are avoided and only a band of frequencies are used. These mid band frequencies are chosen because they generally give a good compromise between robustness and visibility of the watermark.

5. Compute the inverse FFT using the phase components and the modulated magnitude components.

6. Compute the inverse of the perceptual mapping function of step 2. For Weber-Fechner law mapping, the inverse function is an exponential.

7. Replace each watermarked block in the image to obtain the stego image.

8. If the image is a color image, then rescale the red, green and blue components by the relative change in luminance introduced by embedding a watermark. Typically, the red, green and blue pixels occupy a byte each in program memory. If overflow or underflow occurs then the pixel is set to the upper bound 255 or lower bound 0 respectively.

In addition, when selecting the components to be modulated, care must be taken to preserve the symmetry imposed on the Fourier components $F(k_1, k_2)$ by the fact that the image block is real valued: $F(k_1, k_2) = F^*(N_1 - k_1, N_2 - k_2)$, where N_1, N_2 designate the size of the image block.

5.4. Extracting a Watermark

The scaling property of the Fourier Transform allows one to recover the embedded watermark using a different block size to that used when embedding the watermark. This feature is potentially useful for dealing with cropped stego-images. For example, if a single watermark component is embedded at a frequency of 30 cycles/block with a block size of 512 then this will appear as a frequency of 15 cycles/block with a block size of 256. In each case the *normalised* frequency is 15/256.

There is a tradeoff here. Some frequencies cannot be clearly detected using this scheme. For example, a watermark component embedded at a frequency of 29 cycles/block with a block size of 512 will appear at a frequency of 14.5 cycles/block with a block size of 256. This frequency bin does not exist when one uses the Discrete Fourier Transform. This effect is an unfortunate and unavoidable efect of using smaller blocks.

6. THE TEMPLATE

Until now, we only discussed the possibility of using a purely invariant representation in which convey the watermarking information and hence produce a digital watermark which is robust to a given set of transformations. In practice however, this has proven to be difficult to implement so the watermarking scheme proposed in Figure 6 is more interesting from a theoretical than a practical point of view.

In this section, we explore the possibility of abandoning a purely invariant representation and using instead a partially invariant representation. To be more specific, we shall embed a digital watermark using the Fourier Transform magnitude spectrum meaning that the digital watermark is is invariant to cropping and translation. The parameters of the linear transformation suffered by a stego image will be determined by an exhaustive search.

In general, particularly if one is using a spread spectrum technique, one cannot simply serach for the watermark signal unless one already knows the embedded message. For this reason, we shall search for a second auxiliary signal which conveys no information except to act as a reference. This signal we shall call *the template*.

As we saw earlier, in principle if we can work out the linear transformation suffered by an image then we should be able to reverse it. The only constraint is that the transformation is non-singular. We propose to determine the most likely linear transformation suffered by the template This idea is very similar to the idea independently developed by Fleet and Heeger.[26] In their paper, a pattern of eight dots is embedded in the magnitude spectrum of a digital image. The artefacts introduced on the image are not visible. Despite subsequent printing and rescanning operations (which invariably involve some rotation and rescaling) the authors show that it is possible to reliably extract the template and determine the true orientation and scale of the image. The essential difference between the template of Fleet and Heeger[26] and our own is that we generate a random template using a random key.

It could be argued that introducing a template makes the watermarking scheme more vulnerable than before because an attacker may attack the template instead of attacking the digital watermark. In fact, destroying the template is analogous to destroying the watermark and the same arguments apply in each case. Using robust pattern matching techniques it is possible to extract the template even in the presence of a relatively large amount of noise. This noise may be inserted deliberately by an attacker or be inserted inadvertently by image processing such as low pass filtering or lossy image compression. In either case it is very unlikely that the noise interference can disable the template without causing unacceptable damage to image quality first.

There is a good case for arguing that the template is more robust than the watermark. The reason for this is that the template actually has to carry less information than the watermark. For example, suppose it is only possible to recover the watermark if the rotation angle and scaling factor recovered using the template are both 99.9% accurate. To give 99.9% accuracy one needs $\log_2(1000) \approx 9$ bits. Specifying both the angle and the scaling factor therefore needs around 18 bits which is considerably less than the amount of information contained in a typical watermark (about 100 bits).

A very informed and sophisticated attacker can use the key information which was used to generate the template. This attack will be highly effective for disabling public watermarks but cannot work for private and detection watermarks where the key information is kept secret.

Finally, there is the simple question as to what constitutes a "good" template. The one vital property that any template must have is that a template must be unambiguous in the sense that it can never matches well with a translated version of itself. A two dimensional random pattern of points is a serviceable template. Better templates could probably be designed using number theoretic approaches.

In subsections 6.1 and 6.1 we shall describe two simple algorithms for compensating for rotation and scale and proportion and scale.

6.1. Compensating for Rotation and Scale Transformations

In this section, we describe the steps required to locate the template given the follpowing scenario: An image is watermarked, it is then rotated, cropped and scaled. Given only the transformed cropped image how do we determine the intervening distortions? The search technique is as follows:

1. Compute the magnitude spectrum of the image.
2. Transform the magnitude spectrum using a log polar map. Changes in the stego image due to rotations and scales become translations which is very convenient for template matching.
3. Compute the log polar map of the template.
4. Use normalised cross correlation or some other template matching technique[27] to determine the offset.
5. Resynchronise the log polar map of the magnitude spectrum with the template.

At this stage, one could rotate and scale the image back to its original position and then extract the watermark. In fact, it is simpler to extract the watermark directly from the log polar map. This follows from the fact that, since the positions of the points containing watermark information in the magnitude spectrum are known then the positions of these points in the log polar map are also known.

Note that the most efficient way to compute the corelation between two images is to use the FFT. Ironically, the algorithm we have just described to extract a digital watermark bears a very strong resemblance to the completely invariant approach illustrated in Figure 6.

6.2. Compensating for Changes of Proportion and Scale Transformations

Determining the change of proportion (i.e. aspect ratio) suffered by a stego image is trivial. In fact all one needs to do is to exchange the log polar mapping in subsection 6.1 above for a log-log mapping.

Figure 7. *A watermarked image of Lena.*

Figure 8. *A watermarked image of Lena compressed at 15% quality factor. The compression ratio is 100:1.*

7. ROBUSTNESS OF THE WATERMARK

Figure 7 shows an image that was quite strongly watermarked using the techniques described in this paper. Figure 8 shows this image after JPEG compression was applied at 15% quality factor where it is found that that the storage required is less than 1% of that needed for the original image. Not surprisingly, the quality is very low and the image is of little commercial use. In spite of this a 100 bit watermark "The watermark" (in ASCII code) can be recovered from the JPEG compressed image.

Specialised watermark removal algorithms have been proposed in the literature. Stirmark[28] uses both contrast based attacks and geometric attacks. The geometric attacks are not directly addressed using the method proposed in this paper. However, we can give one good example of the robustness of the spead spectrum watermark described in this paper to a contrast based attack. The results are as follows:

- stirmark -i0 -o0 -d128 :

Figure 9. *The watermarked image of Lena after being attacked using "Stirmark"*

- Decoded watermark: The watermark
- Agreement with original = 100.00 percent

stirmark -i0 -o0 -d256 :

- Decoded watermark: The wAtereak
- Agreement with original = 96.15 percent

The stirmark distorted image is shown in Figure 9. It has obviously been very severely distorted. It was surprising even to the authors that this distortion only destroyed 4% of the bits. Other attacks such as Unzign[29] had no effect on the mark.

8. CONCLUSION

In summary, although many of the approaches described in the literature have some aspects in common with the approach described in this paper, such as the use of spread spectrum techniques, the new method brings a number of major improvements. Compared with the few other approaches that are possibly based on invariants, such as that used by Digimarc Corporation in the Picturemarc Technology, to the best of our knowledge, the following aspects are new, non-trivial, and critical:

- There is a major modification to the spread spectrum technique described by Cox el al.[2] Unlike Cox et al.'s approach our system does not require a database of all watermarks that were ever embedded in every image anywhere to extract a message and instead relies on a key.
- We presented a new method for embedding information in an invariant domain by combining a Fourier Transform with a log polar map. The technique of embedding a signal in frequencies in the Fourier Transform domain has been described by many authors.[3,9] These frequencies have previously been chosen either to be robust to additive noise[3] or be oblivious to the original image.[9] The approach used in this paper withstands rotation and scaling by searching for an invisible template in the image magnitude spectrum.

Current work is being directed towards the application of the techniques described in this paper to digital video watermarking.

ACKNOWLEDGMENTS

We wish to thank Dr David McG. Squire, Sergei Starchik, Prof Thierry Pun and Dr Feng-Lin for their extremely helpful advice on the theory of invariants. We also wish to thank Dr A. Z. Tirkel for many stimulating conversations and for exchanging many ideas, particularly regarding spread spectrum techniques. We are grateful to Dr Alexander Herrigel and Adrian Perrig for their work on the security architecture for the digital watermark. We also commend Jean Francois Buisson for his excellent work in implementing the digital watermarking algorithm and the associated security architecture as a Java application.

REFERENCES

1. J. J. K. Ó Ruanaidh, W. J. Dowling, and F. M. Boland, "Watermarking digital images for copyright protection," *IEE Proceedings on Vision, Image and Signal Processing* **143**, pp. 250–256, August 1996. Invited paper, based on the paper of the same title at the IEE Conference on Image Processing and Its Applications, Edinburgh, July 1995.
2. I. Cox, J. Killian, T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for images, audio and video," in *Proceedings of the IEEE Int. Conf. on Image Processing ICIP-96*, pp. 243–246, (Lausanne, Switzerland), September 16-19 1996.
3. J. J. K. Ó Ruanaidh, W. J. Dowling, and F. M. Boland, "Phase watermarking of digital images," in *Proceedings of the IEEE Int. Conf. on Image Processing ICIP-96*, pp. 239–242, (Lausanne, Switzerland), September 16-19 1996.
4. K. Matsui and K. Tanaka, "Video-Steganography: How to secretly embed a signature in a picture," in *IMA Intellectual Property Project Proceedings*, pp. 187–206, January 1994.
5. J. Puate and F. Jordan, "Using fractal compression scheme to embed a digital signature into an image," in *Proceedings of SPIE Photonics East'96 Symposium*, November 1996.
6. I. Pitas, "A method for signature casting on digital images," in *Proceedings of the IEEE Int. Conf. on Image Processing ICIP-96*, pp. 215–218, (Lausanne, Switzerland), September 16-19 1996.
7. M. D. Swanson, B. Zhu, and A. Tewfik, "Transparent robust image watermarking," in *Proceedings of the IEEE Int. Conf. on Image Processing ICIP-96*, pp. 211–214, (Lausanne, Switzerland), September 16-19 1996.
8. J. Delaigle, C. De Vleeschouwer, and B. Macq, "Digital Watermarking," in *Conference 2659 - Optical Security and Counterfeit Deterrence Techniques*, SPIE Electronic Imaging : Science and Technology, (San Jose), Feb. 1996. pp. 99-110.
9. J. Smith and B. Comiskey, "Modulation and information hiding in images," in *Proceedings of the First International Workshop in Information Hiding*, R. Anderson, ed., Lecture Notes in Computer Science, pp. 207–226, Springer Verlag, (Cambridge, UK), May/June 1996.
10. G. Caronni, "Assuring Ownership Rights for Digital Images," in *Reliable IT Systems VIS '95*, H. H. Brueggemann and W. Gerhardt-Haeckl, eds., Vieweg Publishing Company, Germany, 1995.
11. A. Z. Tirkel, G. A. Rankin, R. G. van Schyndel, W. J. Ho, N. R. A. Mee, and C. F. Osborne, "Electronic watermark," in *Dicta-93*, pp. 666–672, (Macquarie University, Sydney), December 1993.
12. A. Z. Tirkel, R. G. van Schyndel, and C. F. Osborne, "A two-dimensional digital watermark," in *ACCV'95*, pp. 378–383, (University of Queensland, Brisbane), December 6-8 1995.
13. B. Schneier, *Applied Cryptography*, Wiley, 2nd ed., 1995.
14. D. M. Squire, *Model-based Neural Networks for Invariant Pattern Recognition.* PhD thesis, Curtin University of Technology, Perth, Western Australia, October 1996.
15. R. Milanese, S. Gil, and T. Pun, "Attentive mechanisms for dynamic and static scene analysis," *Optical Engineering* **34**, pp. 2428–2434, August 1995.
16. J. Gibson, *The Senses Considered as Perceptual Systems*, Houghton-Mifflin, Boston, Massachusetts, 1966.
17. R. L. Pickholtz, D. L. Schilling, and L. B. Milstein, "Theory of spread spectrum communications – a tutorial," *IEEE Transactions on Communications* **COM-30**, pp. 855–884, May 1982.
18. S. Craver, N. Memon, B. Yeo, and M. Yeung, "Can invisible watermarks resolve rightful ownerships?," Tech. Rep. RC20509, IBM Research Division, Yorktown heights, July 1996.
19. I. J. Cox and J.-P. Linnartz, "Public watermarks and resistance to tampering," in *Proceedings of the IEEE Int. Conf. on Image Processing ICIP-97*, vol. 1, (Santa Barbara), October 1997.
20. R. G. van Schyndel, A. Z. Tirkel, and C. F. Osborne, "A digital watermark," in *IEEE Int. Conf. on Image Processing ICIP-95*, pp. 86–90, (Austin, Texas), 1994.
21. R. G. van Schyndel, A. Z. Tirkel, and C. F. Osborne, "Towards a robust digital watermark," in *Dicta-95*, pp. 504–508, (Nanyang Technological University, Singapore), December 5-8 1995.
22. A. Herrigel, J. Ó Ruanaidh, H. Petersen, T. Pun, and S. Pereira, "Copyright protection for digital images based on asymmetric cryptographic and image authentication techniques," in *Electronic Image Capture and Publishing EUROPTO-98*, (Zurich, Switzerland), 1998.
23. R. D. Brandt and F. Lin, "Representations that uniquely characterize images modulo translation, rotation and scaling," *Pattern Recognition Letters* **17**, pp. 1001–1015, August 1996.
24. J. Ó Ruanaidh and T. Pun, "Rotation, translation and scale invariant spread sprectrum digital image watermarking," *Signal Processing* **66**(3), 1998.
25. J. S. Lim, *Two-Dimensional Signal and Image Processing*, Prentice-Hall International, 1990.
26. D. Fleet and D. Heeger, "Embedding invisible information in color images," in *Proceedings of the IEEE Int. Conf. on Image Processing ICIP-97*, vol. 1, pp. 532–535, (Santa Barbara), October 1997.
27. J. P. Secilla, N. Garcia, and J. L. Carrascosa, "Template location in noisy pictures," *Signal Processing* **14**, pp. 347–361, 1988.
28. M. Kuhn, "Stirmark," in *http://www.cl.cam.ac.uk/∼ fapp2/watermarking/image_watermarking/stirmark/*,,
29. "Watermark testing. is your watermark secure?," in *http://www.altern.com/watermark/*,