A Bayesian Approach To Spread Spectrum Watermark Detection and Secure Copyright Protection for Digital Image Libraries

O'Ruanaidh, Joséph John; Csurka, Gabriela Otilia

IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'99), Fort Collins, Colorado, USA, June 23-25, 1999

A Bayesian Approach To Spread Spectrum Watermark Detection and Secure Copyright Protection for Digital Image Libraries

Joseph J. K. Ó Ruanaidh* and Gabriella Csurka

University of Geneva - CUI, 24 rue du Général-Dufour, CH-1211 Geneva 4, Switzerland

oruanaidh@scr.siemens.com, Gabriela.Csurka@cui.unige.ch

Abstract

Digital watermarks have been proposed as a method for discouraging illicit copying and distribution of copyrighted material, and to create secure digital image libraries by adding to images copyright and user-right information. Using a robust digital watermark to detect and trace copyright violations has therefore lot of interest. This paper describes an approach to embedding a digital watermark using the Fourier transform. The paper also addresses the difficult problem of oblivious watermark detection. It is shown that, for the CDMA spread spectrum signal described in the paper, it is still possible to positively detect the presence of a watermark without being able to decode it (and even infer the number of bits contained in the watermark) given only the key used to generate it. Finally, through experimental results the usefulness of such measure is shown.

1 Introduction

Digital media have increasingly taken over and have extended the applications of traditional analog media. Moreover, the popularity of the WWW 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, digital networks and multimedia also afford virtually unprecedented 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.

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) [8, 1], Discrete Fourier Transform (DFT) [9], Wavelets [8], Linear Predictive Coding [7] and Fractals [13]. 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. The term "perceptually significant" is somewhat subjective but it suggests that a good watermark is one which takes account of the behavior 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 [8, 1] to statistical [12] and psycho-visual [15, 3] criteria.

Spread spectrum techniques have become a standard method for encoding information in digital image watermarking. It has several advantageous features such as cryptographic security, robustness against noise, and is capable of achieving error free transmission of the watermark near or at the limits set by Shannon's noisy channel coding theorem [11, 8, 1].

For these reasons the approach presented here uses spread spectrum to encode the message, that can contain information such as the owner of the image, a serial number and perhaps flags which indicate the type of content e.g. religion, pornography, or politics, or alternatively a hash number to a table that contains these informations. System security is based on proprietary knowledge of the keys (or the seeds for pseudo-random generators) which are required to embed, extract or remove an image watermark. The system which provides for the secure exchange of images and keys over the WWW and to ensure a copyright protection to digital image libraries, has been developed and detailed in [6] and will not be addressed here.

To generate the spread spectrum sequence, m-sequences or Gold Codes are used and the encoded

*Current address: Siemens Corporate Research, 755 College Road East, Princeton, NJ 08540, US message is embedded as a digital watermark into the image using the discrete Fourier transform (DFT). Note that there is *no need for the original image* to extract the watermark and to decode the message; hence the term *oblivious* watermarking.

Another contribution of this paper is the use of the Bayesian approach to watermark detection to compute the probability that a watermark was generated with a given key. The most significant feature of this approach is its extreme robustness. All that is required for positive watermark verification is the key. This technique works even when the binary message cannot be decoded. This is interesting, because it suggests that the watermark algorithm can be used either to extract binary messages, or for binding a watermark to the key, i.e. it is not necessary to decode the watermark to verify the ownership.

In order to be able to decode the message even if the image was rotated, scaled or cropped a known template detection is used which allows the inversion of the geometric transformations undergone by the image before extracting the mark; this aspect is not addressed here, see [10].

2 Spread Spectrum

Although spread spectrum systems have been originally developed for military communications, there is a growing interest in the use of this technique in many different fields, including digital image watermarking [1, 9, 16].

The primary advantage of a spread spectrum communication system is its ability to reject undesirable interference caused by either simultaneous transmission of messages through the same channel or by a hostile transmitter attempting to jam the transmission. Another advantage compared to standard modulation techniques such as frequency modulation or pulse code modulation, is that the band spread is independent of the data. Synchronized reception with the code at the receiver is used for despreading and subsequent data recovery.

The basic idea of spread spectrum modulation is that it transforms the narrow band data sequence into a noise-like wide-band signal, using pseudo-random sequences. This can be done, for example, by using phase shift coding [9], by frequency-hop techniques [5] or by the amplitude modulation technique described in section **2.1**.

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 [2]. The obvious solution is to employ *oblivious* watermarks (meaning that an original image is not required to extract the embedded message).

2.1 Encoding the message

Let the message be represented in binary form as $\widehat{\boldsymbol{b}} = (\hat{b}_1, \hat{b}_2, ...\hat{b}_M)^\top$ where $\hat{b}_i \in \{0, 1\}$ and M is the number of bits in the message to be encoded. The binary form of the message $\widehat{\boldsymbol{b}}$ is then transformed to obtain the vector $\boldsymbol{b} = (b_1, b_2, ...b_M)^\top$, with $b_i \in \{1, -1\}$ by exploiting the basic isomorphism between the group[1] ($\oplus$,{0,1}) and the group (*,{1,-1}). The mapping $1 \rightarrow -1$ and $0 \rightarrow 1$ is an extremely important step because it essentially enables us to replace the exclusive-OR operator used in finite field algebra with multiplication. This is useful when decoding real valued sequences such as digital watermarks.

Defining a set of random sequences $\boldsymbol{v}_i$ each corresponding to a bit b_i, the encoded message can be obtained by:

$$\boldsymbol{w} = \sum_{i=1}^{M} b_i \boldsymbol{v}_i = \mathbf{G}\boldsymbol{b} \qquad (1)$$

where $\boldsymbol{b}$ is a $M \times 1$ vector of bits (in ± 1 form), $\boldsymbol{w}$ is a $N \times 1$ vector and $\mathbf{G}$ in $N \times M$ matrix such that the i^{th} column is a pseudo-random vector $\boldsymbol{v}_i$.

Clearly, the effectiveness of this scheme depends on the specific choice for the random vectors $\boldsymbol{v}_i$. Ideally the vectors should be as well separated as possible to get the maximum discrimination between the bits. 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 pseudo-random sequence has direct implications to the reliability and security of the embedded mark. Pseudo-random number generators described in watermarking literature include m-sequences, Gold Codes, Kasami codes, Legendre sequences, perfect maps [4, 16] and also two or higher dimensional arrays [16] in place of the one dimensional pseudo-random vectors.

2.1.1 m-sequences and Gold Codes

It is shown in [11], how pseudo-random sequences can be used to spread the signal spectrum. In order to despread the signal these pseudo-random sequences need to have good randomness properties, long periods and

[1] The bit addition modulo 2, $\oplus$ is equivalent to exclusive-OR.

different sequences must be well separated in term of correlation.

Maximum length sequences or simply m-sequences perfectly fulfil these requirements. They are the largest codes that can be generated by a shift register of a given length. They exist for all integer values n with period $N = 2^n - 1$ and can be easily generated by proper connections of feedback paths in an n-stage shift register circuit [11]. The auto-correlation function and spectral distribution resemble that of white Gaussian noise. Cross-correlation between shifted versions of m-sequences are equal to -1, whereas autocorrelations to the length of the m-sequence. Therefore, one alternative to choose $\boldsymbol{v}_i$ to encode the message $\boldsymbol{b}$ (section **2.1**) is to consider an m-sequence $\boldsymbol{v}_1$ and for each i, $\boldsymbol{v}_{i+1}$ results from $\boldsymbol{v}_i$ by a circular shift of length 1 (each element of $\boldsymbol{v}_i$ is shifted to right and the last element becomes the first).

An other alternative for $\boldsymbol{v}_i$ is to use Gold Codes [11, 4]. A family of Gold Codes is obtained using an m-sequence $\boldsymbol{v}_1$ and a q-decimation of it. The decimation $\boldsymbol{v}'_1$ is obtained by sampling every q^{th} element of $\boldsymbol{v}_1$. Note that $\boldsymbol{v}'_1$ has period N if and only if gcd(N,q)=1, where "gcd" denotes the greatest common divisor. Each element $\boldsymbol{v}_{i+1}$ of the family can then be obtained as follows: take $\boldsymbol{v}'_1$ (of length N), shift it by i (circular shift) and multiply element by element with the vector $\boldsymbol{v}_1$.

Gold sequences have the advantage that for a given register length n there are more choices for the "key" than with shifted m-sequences. Indeed, for a register length n, there are $2^n - 1$ possible m-sequences ($2^n - 1$ possible seed as initial element in the register) but $2^{2n} - 1$ possible family of Gold Codes as there is in addition the possibility to choose a different decimation (a second linear feedback shift register of length n [11, 4]). In addition, it is known that Gold sequences have better cross correlation properties if only part of the sequence is used. This could have implications if the watermark is partially destroyed by image cropping or filtering.

3 DFT Watermark

The watermark, in the form of a spread spectrum sequence, is embedded in a mid-frequency range of the discrete Fourier transform (DFT) domain. The upper and the lower frequency bound are fixed in advance as a compromise between the visibility of the watermark (low frequencies contain the most of the image information) and robustness to lossy compression (which remove high frequencies). Also, during decoding (in oblivious watermarking) the block size is unknown since the image may have been transformed by cropping or other image processing.

The spread spectrum message is added to the magnitudes of the DFT and the phase is left unaltered. From the marked spectrum and the unchanged phase the DFT is inverted (inverse Fourier transform) yielding the watermarked image. The strength of the watermark can be set either interactively or adaptively as a function of the average and standard deviation of the DFT components of the chosen frequency range.

To extract the watermark the magnitude of the DFT domain is considered. Since the mid-frequency range where the watermark was added is known, the sequence $\boldsymbol{w}' = \boldsymbol{w} + \boldsymbol{e}$, where $\boldsymbol{w}$ is the added watermark and the error $\boldsymbol{e}$ contains the image and/or additive noise, can be extracted. To decode the message $\boldsymbol{b}$ from $\boldsymbol{w}'$, the correlation property of the m-sequences and Gold Codes described in section **2.1** are exploited, i.e the cross-correlation between $\boldsymbol{v}_i$ and $\boldsymbol{v}_j$ is -1 and for $i \neq j$ and N for $i = j$.

However, if the modification suffered by the image are geometric transformations such as rotation, scaling, cropping the positions where the mark was embedded also change. In order to be able to synchronize and to decode the message in these cases, a *template* is used to detect and invert the geometric transformations undergone by the image prior to extracting the mark. The applied algorithm is detailed in [10] and will not be addressed here.

4 The Bayesian Approach

The purpose of this approach is to answer the following questions:

1. Consider a binary string extracted from an image, for which almost all bits agree with a known binary sequence message which may have been embedded in the image. Generally, one can expect that about 50% of the bits of a random sequence will agree with the watermark. What is the probability that the almost perfect agreement occurred at random?

2. Given only the key possibly used to generate a watermark, what is the probability that the watermark was generated using that key?

The last question is extremely interesting because it suggests that one can use the watermark algorithm either to extract binary messages or for binding a watermark to the key and to determine the ownership without decoding the watermark.

Being able to detect a watermark without necessarily being able to decode it, is therefore highly useful since it can help to prove ownership (the owner of the given key) in the case when due to noise the message can only partially be decoded. One expects that watermark detection will always be more robust than watermark decoding because in detection one is essentially transmitting a single bit of information which is to say whether a watermark is present or not.

4.1 Known binary message

In order to answer the first question, let us consider the probability that a random sequence will have a certain number of bits in common with our sequence. We can easily show that this probability is given by the Bernoulli distribution:

$$p(i) = \frac{1}{2^N} \left(\begin{array}{c} N \\ i \end{array} \right)$$

where N is the number of bits in both messages and i is the number of bits found to be in common.

The implications of this result are quite far reaching. If one decodes a 100 bit watermark and finds that 80% of the bits are "correct" then one can be fairly sure that the watermark was indeed found. This is because the probability of getting 80% or more bits correct is at random 2.17×10^{-9}. This is the probability of a false alarm – where one would say that a watermark is present when in fact there is none.

4.2 The probability of a watermark given only the key

We now turn our attention to the second case, i.e. given only the key possibly used to generate a watermark what is the probability that a watermark was generated using that key? We confine our interest to the encoding spread spectrum messages described earlier. The watermark $\boldsymbol{w}' = \mathbf{G}\boldsymbol{b} + \boldsymbol{e}$ is a linear combination of pseudo-random sequences corrupted by noise, where $\boldsymbol{e}$ is a noise vector corrupting the watermark (including the image and additive noise).

If we assume that the noise is Gaussian distributed we can apply the Bayesian approach described in [14] and obtain the probability that a spread spectrum signal $\boldsymbol{w}'$ extracted from the image I contains a message of length M encoded with the key k:

$$p\,(k, M \mid \boldsymbol{w}', I) \propto \frac{\pi^{-N/2}\,\Gamma\big(\frac{M}{2}\big)\,\Gamma\big(\frac{N-M}{2}\big)\,\big|\mathbf{G}^\top\mathbf{G}\big|^{-1/2}}{4\,R_\delta\,R_\sigma\,\big(\boldsymbol{b}'^\top\boldsymbol{b}'\big)^{M/2}\,\big(\boldsymbol{w}'^\top\boldsymbol{w}' - \boldsymbol{f}^\top\boldsymbol{f}\big)^{(N-M)/2}} \qquad (2)$$

where

$$\boldsymbol{b}' = \big(\mathbf{G}^\top\mathbf{G}\big)^{-1}\mathbf{G}^\top\boldsymbol{w}' \quad \text{and} \quad \boldsymbol{f} = \mathbf{G}^\top\boldsymbol{b}$$

are the least squares estimate for the bits and the least squares fit respectively. R_σ and R_δ are irrelevant constants introduced as normalization factors.

As we are interested in analyzing spread spectrum signals composed of a linear combination of Gold sequences or m-sequences, we can show that the $M \times M$ matrix $\mathbf{G}^\top\mathbf{G}$ is of the following form:

$$\mathbf{G}^\top\mathbf{G} = \left[\begin{array}{ccccc} N & -1 & -1 & \cdots & -1 \\ -1 & N & -1 & \cdots & -1 \\ \cdots & & & & \\ -1 & -1 & -1 & \cdots & N \end{array} \right] \qquad (3)$$

The two terms in expression (2) which require the most computation are $\mathbf{A} = (\mathbf{G}^\top\mathbf{G})^{-1}$ and $\big|\mathbf{G}^\top\mathbf{G}\big|$. However, both the determinant and the inverse can be computed in closed form:

$$\big|\mathbf{G}^\top\mathbf{G}\big| = (N + M - 1)(N + 1)^{M-1}$$

$$\mathbf{A}(i,j) = \left\{ \begin{array}{ll} \frac{2^N - M + 1}{2^N\,(2^N - M)} & \text{if} \quad i = j \\ \frac{1}{2^N\,(2^N - M)} & \text{if} \quad i \neq j \end{array} \right.$$

In a similar way, we obtain that the probability that no message was embedded (a message of length 0) with a given key in the image is given by:

$$p(k, 0 \mid \boldsymbol{w}', I) \propto \frac{\pi^{-N/2}\Gamma\big(\frac{N}{2}\big)}{2R_\sigma(\boldsymbol{w}'^\top\boldsymbol{w}')^{N/2}} \qquad (4)$$

Finally, to decide if a given key was used or not to generate a watermark, we compute the relative log-probability:

$$log\Big(\frac{p(k, M \mid \boldsymbol{w}', I)}{p(k, 0 \mid \boldsymbol{w}', I)}\Big) \qquad (5)$$

and we compare it to 0.

5 Experimental results

5.1 Synthetic data

For an example of equation (2) at work we consider a synthetic watermark and corrupt it by Gaussian noise. The amplitude of the Gaussian noise is twenty times larger than the watermark amplitude or equivalently the Gaussian noise is 400 times more powerful. Here we shall consider the problem of determining the probable length of a watermark generated using a given key. For this we take the watermark above and we compute the probability that a message of length $m, m \in \{0, 25\}$ was encoded with the key.

The length of the message is 10 bits and we compute the probability that the watermark was generated with the correct key. Figure 1(left) shows the probabilities relative to that of a watermark length of 9 bits. The inferred length is 10 bits which was exactly the value used when generating this synthetic data.

Figure 1 (middle and right) shows what happens when one attempts to infer the length of a watermark when in fact there is none present or when a wrong key was used. The method infers that the watermark is of zero bits in length. Quite logically, as in the first case there is no mark and in the second case the watermark data generated with the wrong key appears as "noise", so it is as if no watermark were present.

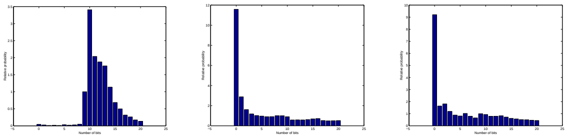

Figure 1: *An attempt to determine the correct number of bits in the watermark given only the key. Left: using the correct key. Middle: using the wrong key. Right: there is no watermark present at all.*

5.2 Real data

We used three real images for our experiences. In each cases we encoded the message “national” (64 bit) with the key (seed) 1967 and we marked the images (Figure 2).

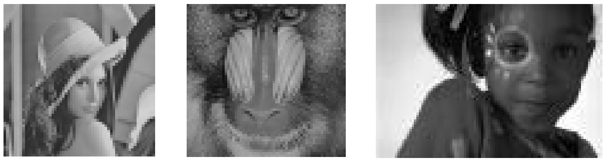

Figure 2: *Three marked images used to test. Up: Lena and Mandrill, Down: Girl.*

Figure 3 shows several attempts on the marked and unmarked images to detect and decode the message using different keys. We can see that the relative log-probabilities were below the threshold (generally between 0 and -4), except, when we used the correct key and the marked images. In these latter cases the values were between 1500-3500 depending on the image. Note, that we have a false alarm, i.e. we obtain a positive value in the case of one of the unmarked image. However, this value is close to 0 (0.46) and it can be avoided by choosing a threshold $t > 0$) (e.g. 1) instead of 0, by accepting to have occasional false rejections (noisy cases).

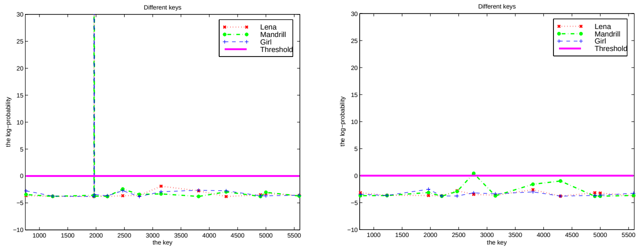

Figure 3: *Attempt to detect watermark given different keys. Left: marked images. Right: unmarked images.*

Furthermore, we applied several rotations of different angles on each images and we compared bit by bit the encoded and the decoded messages. In each case we were able to decode all 64 bits (100%). The corresponding relative log-probabilities are shown in the Figure 4(left) and we can see that each value was above the threshold (either 0 or 1).

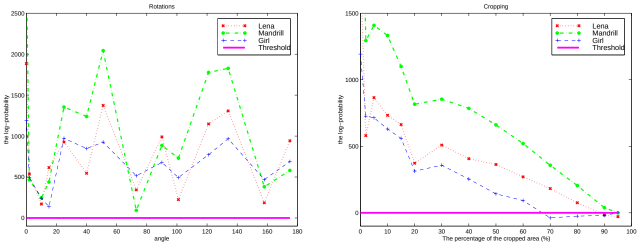

Figure 4: *Left: Results of several rotations. Right: Results of different centered cropping.*

We also tested the resistance of our algorithm and the behavior of the Bayesian approach to cropping. Figure 4(right) shows the obtained relative log-probabilities. We can see that the log-probabilities decrease as the cropped area increases. However, 100% of bits can be retrieved until only 30-40% of the original image data is retained. When a smaller percentage of the image remains, we first begin to lose some bits and finally only 35-65% of bits corresponds to the encoded message[2].

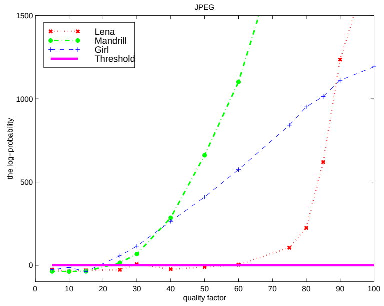

Figure 5: *Results of JPEG lossy data compression.*

Finally we attempted to detect and decode the message from marked images after JPEG lossy data compression (Figure 5). The log-probabilities increase with the quality factor. 84-100% of bits were decoded for a quality factor over 25 (100% for a quality factor $\geq$ 40) excepting for Lena.

6 Conclusion

We have presented here an approach for image watermarking in which the watermark is embedded in the DFT domain. The message is encoded in the spread spectrum signal using an owner key to ensure the security. The properties of the m-sequences or Gold Codes

[2] The percentage of bits that will agree between a random sequence and the encoded message (section **4.1**).

were studied and exploited in order to generate the spread spectrum sequence. Furthermore, a Bayesian approach was developed that acts concurrently to the message decoding, by estimating the probability that a watermark was generated given only the key. We can therefore use the watermark algorithm either to extract binary messages or to bind the watermark to the key so the *actual message is irrelevant*.

Experimental results show on one hand the efficiency and the robustness of our watermark detection algorithm and, on the other hand, the behavior and the usefulness of the added Bayesian measure.

The most significant feature of the presented approach is its extreme robustness. This justifies its integration into a watermarked-based copyright protection system for digital libraries, which is an acknowledged and pressing need.

Acknowledgments

We are grateful to T. Mittelholzer, Dr A. Herrigel and Digital Copyright Technologies Switzerland for the ongoing collaboration and the several interesting discussion we had concerning the pseudo-random sequences and watermarking techniques. This work is financed by the European Esprit Open Microprocessor Initiative (project JEDI-FIRE) and by the Swiss Priority Program on Information and Communication Structures (project Krypict) and is part of the European Patent application EU978107084.

References

[1] I. Cox, J. Killian, T. Leighton, and T. Shamoon. Secure spread spectrum watermarking for images, audio and video. In *Proc. of the IEEE Int. Conf. on Image Processing*, pages 243–246, Lausanne, Switzerland, September 1996.

[2] S. Craver, N. Memon, B. Yeo, and M. Yeung. Can invisible marks resolve rightful ownerships? In *IS&T/SPIE Electronic Imaging '97: Storage and Retrieval of Image and Video Databases*, 1997.

[3] J. F. Delaigle, C. De Vleeschouwer, and B. Macq. Digital Watermarking. In *Conference 2659 - Optical Security and Counterfeit Deterrence Techniques*, San Jose, February 1996. SPIE Electronic Imaging : Science and Technology. pp. 99-110.

[4] E. H. Dinan and B. Jabbari. Spreading codes for direct sequence CDMA and wideband CDMA cellular network. *IEEE Communications Magazine*, June 1998.

[5] S. Haykin. *Communications Systems.* Wiley, 3rd edition, 1994.

[6] Alexander Herrigel, Joe J. K. Ó Ruanaidh, H. Petersen, S. Pereira, and T. Pun. Secure copyright protection techniques for digital images. In *International Workshop on Information Hiding*, Portland, OR, USA, April 1998.

[7] K. Matsui and K. Tanaka. Video-Steganography: How to secretly embed a signature in a picture. In *IMA Intellectual Property Project Proceedings*, pages 187–206, January 1994.

[8] Joe J. K. Ó Ruanaidh, W. J. Dowling, and F. M. Boland. Watermarking digital images for copyright protection. *IEE Proceedings on Vision, Signal and Image Processing*, 143(4):250–256, August 1996.

[9] Joe J. K. Ó Ruanaidh and T. Pun. Rotation, scale and translation invariant spread spectrum digital image watermarking. *Signal Processing*, 66(3):303–317, May 1998. (Special Issue on Copyright Protection and Control, B. Macq and I. Pitas, eds.).

[10] S. Pereira, J. J. K. Ó Ruanaidh, F. Deguillaume, G. Csurka, and T. Pun. Template based recovery of Fourier-based watermarks using Log-polar and Log-log maps. IEEE Int. Conf. on Multimedia Computing and Systems, Special Session on Multimedia Data Security and Watermarking, Florence, Italy, June 1999.

[11] R. L. Pickholtz, D. L. Schilling, and L. B. Milstein. Theory of spread spectrum communications – A tutorial. *IEEE Transactions on Communications*, COM-30(5):855–884, May 1982.

[12] I Pitas. A method for signature casting on digital images. In *Proc. of the IEEE Int. Conf. on Image Processing*, pages 215–218, Lausanne, Switzerland, September 1996.

[13] 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.

[14] J. J. K. Ó Ruanaidh and W. J. Fitzgerald. *Numerical Bayesian Methods Applied to Signal Processing.* Series on Statistics and Computing. Springer-Verlag, 1996.

[15] M. D. Swanson, B. Zhu, and A. Tewfik. Transparent robust image watermarking. In *Proceedings of the IEEE Int. Conf. on Image Processing ICIP-96*, pages 211–214, Lausanne, Switzerland, September 1996.

[16] A. Z. Tirkel, C.F. Osborne, and T.E. Hall. Image and watermark registration. *Signal processing*, 66:373–383, 1998.