PHASE WATERMARKING OF DIGITAL IMAGES

J.J.K. Ó Ruanaidh, W.J. Dowling and F.M. Boland

Department of Electronic and Electrical Engineering
University of Dublin, Trinity College
Dublin 2
Ireland
E-mail: oruanaidh@ee.tcd.ie

ABSTRACT

A watermark is an invisible mark placed on an image that can be detected when the image is compared with the original. This mark is designed to identify both the source of an image as well as its intended recipient. The mark should be tolerant to reasonable quality lossy compression of the image using transform coding or vector quantization. Standard image processing operations such as low pass filtering, cropping, translation and rescaling should not remove the mark. Spread spectrum communication techniques and matrix transformations can be used together to design watermarks that are robust to tampering and are visually imperceptible. This paper discusses techniques for embedding such marks in grey scale digital images. It also proposes a novel phase based method of conveying the watermark information. In addition, the use of optimal detectors for watermark identification is also proposed.

1. WATERMARKING

Zhao and Koch [1] investigated an approach to watermarking images based on the JPEG image compression algorithm. Their approach is to segment the image into individual 8×8 blocks. Only eight coefficients occupying particular positions in the 8×8 block of DCT coefficients can be marked. These comprise the low frequency components of the image block, but exclude the mean value coefficient (at coordinate (0,0)) as well as the low frequencies at coordinates (0,1) and (1,0). Three of these coefficients are selected using a pseudo-random number generator to convey information.

Tirkel et al. [2, 3] and van Schyndel et al. [4, 5] have applied the properties of m-sequences to produce watermarks that are resistant to filtering, image cropping and are reasonably robust to cryptographic attack. The original image is not required to decode the mark. Recent work [5] indicates progress towards producing more robust watermarks.

Matsui and Tanaka [6] have applied linear predictive coding for watermarking video, facsimile, dithered binary pictures and colour and grey scale images. Their approach to hiding a watermark is to make the watermark resemble quantization noise. To a certain extent, their approach can be considered to be perceptually adaptive in that quantization noise is concentrated around edges and textured features. Cox et al. [7] believe that this method may not be robust to cropping.

Ó Ruanaidh et al. [8], and Cox et al. [7] independently developed perceptually adaptive transform domain based methods for watermarking. In contrast to the previous approaches listed above, the emphasis was on embedding the watermark in the *most significant* components of an image. The general approach used in these papers is to divide the image into blocks. Each block is mapped into the transform domain using either the Discrete Cosine Transform [9, 10, 11], the Hadamard Transform [9] or the Daubechy Wavelet Transform [11]. The bits are placed by incrementing selected coefficients to encode a '1' and decrementing to encode a '0'. Cox et al. [7] embedded Gaussian sequences to produce robust, continuous valued watermarks that are resistant to tampering. Only the components that are most significant to image integrity are marked.

2. THE DFT

The algorithm described here depends on the properties of the Discrete Fourier Transform

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

The inverse transform is

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

where $\beta = (N_1 N_2)^{-1/2}$.

The Discrete Fourier Transform of a real image is generally complex valued. This leads to a magnitude and phase representation for the image. Hayes [12] investigated the relative importance of the magnitude and phase components of the DFT and their effect on the intelligibility of an image. It is demonstrated quite conclusively that the phase is more important than the magnitude of the the DFT values.

This is interesting from our point of view for several reasons. First, a watermark that is embedded in the phase

This work was supported by Forbairt.

of the DFT would be quite robust to tampering. The core information contained in watermarks is almost always encoded with a high degree of redundancy. Therefore background clutter and phase distortions deliberately introduced by an "enemy" to impede transmission of the watermark would have to be noticeably large in order to be successful. This would cause unacceptable damage to the quality of the image. Second, from communications theory, it is well known that angle modulation can possess superior noise immunity when compared to amplitude modulation. We also find that phase based watermarking is relatively robust to changes in image contrast.

2.1. Phase watermarking

The condition that the image be real imposes the following constraint on the values of the DFT:

$$F(k_1, k_2) = F^*(N_1 - k_1, N_2 - k_2) \tag{3}$$

Therefore, when watermarking, changes of phase must preserve negative symmetry:

$$\begin{array}{lcl} \angle F(k_1, k_2) & \leftarrow & \angle F(k_1, k_2) + \delta \\ \angle F(N_1 - k_1, N_2 - k_2) & \leftarrow & \angle F(N_1 - k_1, N_2 - k_2) - \delta \end{array} \tag{4}$$

Similarly, changes of magnitude must preserve positive symmetry.
A DFT value $F(k_1, k_2)$ is marked if

$$|F(k_1, k_2)|^2 / \sum_{r_1=1}^{N_1-1} \sum_{r_2=1}^{N_2-1} |F(r_1, r_2)|^2 > \epsilon \tag{5}$$

where ϵ is a constant. The only exceptions are that $F(0,0)$ (the so called DC component) as well as $F(N_1/2, N_2/2)$ cannot be changed if equation 3 is to hold.

3. EFFECT OF ADDITIVE GAUSSIAN NOISE

In this section, we outline a derivation for the probability density function of the angle of a phasor of length r as it is subjected to two dimensional additive white Gaussian noise of standard deviation σ.

Referring to Figure 1 and denoting $|OP|$ by r, $|OQ|$ by l and angle $\angle QOP$ by α we may write:

$$l \cos\alpha = e_1 + r \tag{6}$$

$$l \sin\alpha = e_2 \tag{7}$$

where e_1 and e_2 are independent Gaussian variates of standard deviation σ. The joint density $p(e_1, e_2)$ is given by:

$$p(e_1, e_2) = \frac{1}{2\pi\sigma^2} \exp\left[-\frac{e_1^2 + e_2^2}{2\sigma^2}\right] \tag{8}$$

Substituting expressions 6 and 7 it is easy to show that the joint density for the random variables l and α is

$$p(l, \alpha) = \frac{l}{2\pi\sigma^2} \exp\left[-\frac{\left(r^2 - 2rl\cos\alpha + l^2\right)}{2\sigma^2}\right] \tag{9}$$

The joint density $p(l, \alpha)$ can be integrated with respect to l to give the marginal density $p(\alpha)$. If $r >> \sigma$ the integral may be well approximated using techniques described by Ó Ruanaidh and Fitzgerald [13] to obtain the following approximation:

$$p(\alpha) = \frac{1}{\sqrt{2\pi(\sigma/r)^2}} \exp\left[-\frac{\sin^2\alpha}{2(\sigma/r)^2}\right] \tag{10}$$

For small values of α we have $\sin\alpha \approx \alpha$ which gives a Gaussian distribution:

$$p(\alpha) = \frac{1}{\sqrt{2\pi(\sigma/r)^2}} \exp\left[-\frac{\alpha^2}{2(\sigma/r)^2}\right] \tag{11}$$

The standard deviation of the phase distribution is σ/r which shows that the distortion effects of additive noise are greatly diminished if only those DFT components which have the largest DFT magnitudes are marked.

The performance of the approximations in expressions 10 and 11 are assessed in Figure 2. The Gaussian approximation is very good in the body of the distribution and it is only slightly outperformed by expression 10 in the tails.

4. RECOVERY OF THE MARK

There are two distinct methods for recovering the watermark from a marked image. The first, and most obvious method, is to simply compare the marked image with an unmarked original. The second method which does not require any comparison is less obvious. If the information bearing quantities are preprocessed and quantized prior to marking then deviations from these quantized states could be used to convey information. Matsui and Tanaka [6] present an example that uses this idea. Coarse quantization allows deeper marking which in turn gives improved robustness of the mark. However this is at the cost of a reduction in information content and image degradation. The optimal trade off between these various factors is heavily dependent on psychovisual considerations.

Optimal statistical techniques will have a very important rôle to play in authenticating watermarks. In the Cox et al. [7] method of watermarking, the Gaussian sequence recovered from a watermarked image is compared with candidates stored in a database of watermarks. A match is

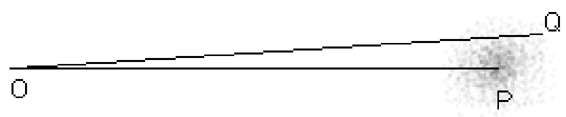

Figure 1: This figure shows the effect of Gaussian noise. Noise added to the vector $|OP|$ gives rise to vector $|OQ|$. This results in an angular displacement (or phase distortion) given by the angle QOP.

obtained by selecting the one which gives the largest correlation coefficient. This is a simple but reasonably effective technique.

To examine the effect of using an optimal detector similar results to those obtained by Cox et al. [7] in figure 5 of their paper were generated. The data were analysed using Bayesian techniques [13]. The odds against the watermark being one of the 999 other possibilities came out as a resounding $10^{15}:1$! Quoting odds to support the authenticity a watermark would be a more satisfactory way of presenting evidence in the case of a dispute that goes to court. An optimal detector is only as good as the model assumptions on which it is based. Therefore it would be beneficial to carry out further more detailed studies on the nature of the distortions suffered by a watermark.

5. RESULTS

A standard image was watermarked using the DFT. Figure 3 shows the original grey-scale image of 256×256 pixels. A block size of 8×8 pixels was used by the watermarking algorithm. The watermarked image is shown in Figure 4. Only the DFT phase is used to embed the codeword. A total of 9920 bits is embedded in Figure 4. Despite the presence of the watermark the picture does not contain any visible artifacts.

Figure 5 shows the absolute difference between the original image and the marked image scaled by a factor of 64. It is interesting to note that the biggest difference occurs around the edges.

In experiments the watermarked image was compressed using a JPEG encoder. At present, it is possible for the watermark to survive 15:1 compression. This is likely to be improved upon with further work.

6. CONCLUSION

In this paper, we argued that watermarking needs to be *adaptive* in order to be robust. In direct contrast to many other techniques, with the notable exception of Cox et al. [7], the method presented here places the watermark on the *most significant* frequency components of an image. The logic behind the premise is quite simple. A watermark that is non-intrusive is one which resembles the image it is designed to protect. By virtue of its similarity to the image, any operation that is intentionally performed to damage the watermark in some way also will unavoidably damage the image. This idea is consistent with the use of phase in the transform domain as a method of conveying information because phase information is more important to the viewer than magnitude information.

The use of optimal statistical detectors for the identification and substantiation of watermarks was also proposed. The use of such detectors will place the difficult area of authenticating watermarks on a far more objective footing. In addition it will facilitate the detection of very weak watermarks which is beyond the capability of simple detectors.

Future work will concentrate on integrating aspects of the Human Visual System into watermarking algorithms. In addition, a detailed study of the effects of image distortion on a watermark will be undertaken with a view to improving watermark detection. Finally, novel techniques will be devised to make it possible to detect a watermark without requiring the original unmarked image.

7. REFERENCES

[1] J. Zhao and E. Koch. Embedding robust labels into images for copyright protection. Technical report, Fraunhofer Institute for Computer Graphics, Darmstadt, Germany, 1994.

[2] A. Z. Tirkel, G. A. Rankin, R. G. van Schyndel, W. J. Ho, N. R. A. Mee, and C. F. Osborne. Electronic water mark. In *Dicta-93*, pages 666–672, 1993.

[3] A. Z. Tirkel, R. G. van Schyndel, and C. F. Osborne. A two-dimensional digital watermark. In *ACCV*, Singapore, 1995.

[4] R. G. van Schyndel, A. Z. Tirkel, and C. F. Osborne. A digital watermark. In *IEEE International Conference on Image Processing*, pages 86–90, Austin, Texas, 1994.

[5] R. G. van Schyndel, A. Z. Tirkel, and C. F. Osborne. Towards a robust digital watermark. In *Dicta-95*, 1995.

[6] 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.

[7] I. Cox, J. Killian, T. Leighton, and T. Shamoon. Secure spread spectrum communication for multimedia. Technical report, N.E.C. Research Institute, 1995. ftp://ftp.nj.nec.com/ pub/ingemar/ papers/watermark.ps.Z.

[8] 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*, 1996.

[9] R. J. Clarke. *Transform Coding of Images.* Academic Press, London, 1985.

[10] W. B. Pennebaker and J. L. Mitchell. *JPEG Still Image Compression Standard.* Van Nostrand Reinhold, New York, 1993.

[11] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes in C.* Cambridge University Press, second edition, 1992.

[12] M. H. Hayes. The reconstruction of a multidimensional sequence from the phase or magnitude of the FFT. *IEEE Trans on Acoustics, Speech and Signal Processing*, pages 140–154, April 1992.

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

Figure 2: This plot compares the theoretical phase distribution with an experimental phase distribution when $r = 5\sigma$. The histogram was computed using 100000 samples. The dotted curve and the solid curve correspond to the theoretical distributions in expression 10 and expression 11 respectively.

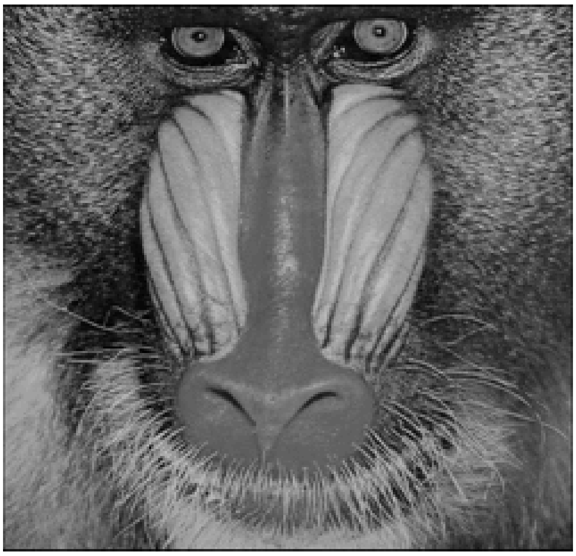

Figure 3: Original grey-scale image of 256 $\times$ 256 pixels. (8 bits/pixel)

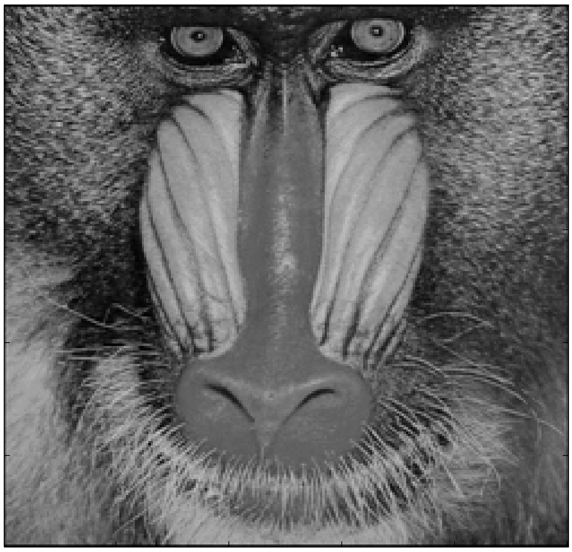

Figure 4: Watermarked grey-scale image using the FFT.

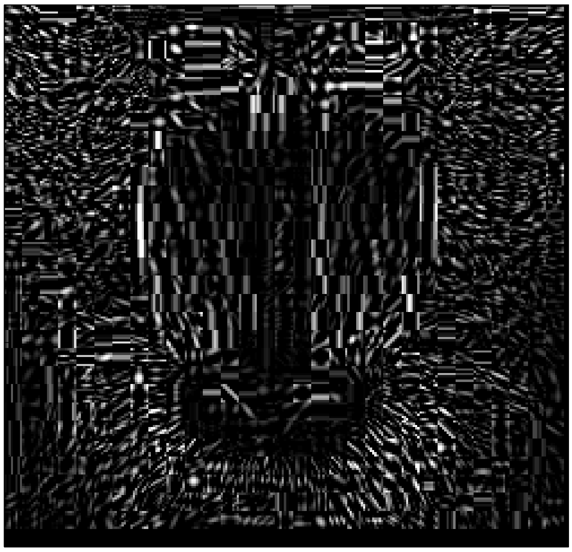

Figure 5: Absolute difference between watermarked and original grey-scale image scaled by a factor of 64.