Watermarking Digital Images for Copyright Protection

F.M. Boland[†], J.J.K. Ó Ruanaidh[†] and C. Dautzenberg[‡]

[†]Trinity College Dublin, Ireland

[‡]Rheinisch-Westfaelische Technische Hochschule, Aachen, Germany

Abstract

A watermark is an invisible mark placed on an image that can only be detected when the image is compared with the original. This mark is designed to identify both the source of a document as well as its intended recipient. This paper discusses various techniques for embedding such marks in grey scale and colour digital images.

1 INTRODUCTION

Computers, printers and high rate transmission facilities are becoming less expensive and more generally available. It is now feasible and very economical to transmit images and video sequences using computer networks rather than sending hard copies by post. In addition, images may be stored in databases in digital form. A major impediment to the use of electronic distribution and storage is the ease of intercepting, copying and redistributing electronic images and documents in their *exact* original form. As a result, publishers are extremely reluctant to use this means of disseminating material.

The commercial possibilities for the World Wide Web are steadily becoming more appreciated. However it is clear that in order for these possibilities to be realized that an integrated approach for the secure handling, issue and duplication of issued documents is required.

Brassil et al. [1] have investigated different methods for marking text within documents with a unique binary code word which serves to identify legitimate users of the document. The code word is embedded in the document by making subtle modifications to the structure of a document such as modulation of line width and interword spacing as well as modification of character fonts. The presence of the code word does not visibly degrade the document, but can be readily detected by making a comparison with the original. Standard document handing operations such as photocopying and scanning do not remove the mark.

The same idea may be extended to include the protection of images. In this paper, we begin by specifying the requirements that an effective image watermarking scheme must possess. A review of current techniques is presented and novel techniques based on image transforms are then described.

1.1 REQUIREMENTS FOR WATERMARKING ALGORITHM

The work in this paper examines strategies for the watermark to meet the following criteria:

- The image must not be visibly degraded by the presence of the mark while at the same time a unique identifier with high information content is produced.
- The mark must be readily recoverable by some form of comparison with the original image.
- The mark must be strongly resistant to detection and decoding without access to the original. It must be strongly resistant to attack and it should cause a significant loss of image quality for it to be destroyed. In addition, the mark must be tolerant to reasonable quality lossy compression of the image.

1.2 DIGITAL COMMUNICATIONS

The task of embedding a watermark in an image and detecting and decoding the mark may be regarded as a problem in digital communications. There are three components [2] in the solution of this problem:

- Forms of transmission pulse must be identified that can transmit information reliably and yet introduce no artifacts visible even to a very careful observer.
- Digital signal modulation techniques are required to place the desired information onto the transmitted pulses.
- Innovative error-control coding and digital signature techniques are required to ensure reliable and secure communication of the mark as well as authentication of the encoded message.

It will be assumed without loss of generality that the mark is encoded in the form of a binary bit string.

The factors affecting the choice of form of transmission pulses are quite complex. First, there is the need for robustness. Any operation that may be carried out on the image can degrade transmission of the watermark. The second factor is visibility. Intuitively, one can see that less information can be hidden on flat featureless regions of the image. It should be possible to incorporate more information into those parts of the image that contain more texture or around edges. Psychovisual phenomena are obviously factors in the transmission of hidden information.

Kurak and McHugh [3] have considered the possible application of redundant features in an image to the transmission of information. Their concern was the transmission of dangerous viruses (or "Trojan horse programs") in the low order bits of a data stream. They note that merely viewing an image is not sufficient for detecting the presence of some form of corruption. Depending on the texture of the image and the quality of a computer monitor it is possible

Image Processing And Its Applications, 4–6 July 1995
Conference Publication No. 410, © IEE 1995.

to exploit the limited dynamic range of the human eye to hide low quality images within other images. Walton [4] has developed a technique for introducing checksums in the low order bits of an image to prevent unauthorized tampering. Dautzenberg and Boland [5] examined using the low order bits as a possible part of a scheme for introducing watermarks into images. This approach gave very poor results because standard lossy compression schemes, such as JPEG [6], tend to have the effect of randomizing the low order bits during the quantization stage of image compression.

2 THE BLOCK MEAN APPROACH

Dautzenberg and Boland [5] and Caronni [7] have investigated another simple technique for embedding watermarks in images. An image may be divided up into blocks. The mean of each block may then be incremented to encode a '1' or decremented to encode a '0' (or vice versa). This is termed bi-directional coding. Alternatively, the mean may be incremented to encode a '1' and be left untouched to encode a '0'. This is termed unidirectional coding. Of the two forms of coding bi-directional coding is the more robust but unidirectional coding has the advantage that it is possible to arrange matters that a watermark can be detected without comparing the image to the original. Dautzenberg and Boland examined two different approaches for placing the blocks inside images:

Chessboard pattern: The blocks are arranged side by side to tile over the entire image.

Blocks with borders pattern: The blocks are arranged side by side, but are surrounded by a border which is not marked.

The block-mean approach suffers from the grave disadvantage that an enemy that is in possession of a number of independent copies of the image can compare the different copies and read most, if not all, of the encoded message. Caronni shows that the expected number of undetected bits decreases exponentially with the number of copies. Caronni combats this particular weakness by randomizing both the size of the blocks as well as the positions of the blocks inside the image.

Despite its simplicity, the block-mean method of marking images has proven to be highly robust to lossy image compression, photocopying and colour scanning and dithering.

The number of bits that may be encoded using the block-mean approach equals the number of blocks, and this in turn depends on the size of the image and the block size, as well as the width of borders around blocks. Realistically, the number of bits that one can expect to encode is in the order of one hundred bits. This capacity may be adequate for some applications, even after taking into account the need for redundancy in the code for error detection and correction as well as code word authentication. However, this capacity is quite tiny in comparison with the storage required for the image.

3 THE WATERMARKING ALGORITHM

It is possible to achieve much higher storage capacities using image transform coding techniques [8]. Candidate image transforms are based on standard image compression techniques and include the use of the Discrete Cosine Transform [6, 9], Wavelet Transforms [10], Walsh-Hadamard Transform [8, 11] and the Fast Fourier Transform.

3.1 THE ALGORITHM

This subsection describes the watermarking algorithm. First, a simple form of modulation for placing bits on an image is outlined. Second, a technique for determining the number of bits to be placed at given locations in the image is also described.

3.1.1 Amplitude Modulation The following algorithm, which is a hybrid between amplitude modulation and frequency shift keying has been applied to watermarking:

1. Divide image into blocks.
2. Subtract the mean of the block from each pixel in the block.
3. Normalize pixel values within each block so that they range between -127 and 127.
4. Carry out transform on image block.
5. Modulate selected coefficients of the transformation (e.g.using bi-directional coding).
6. Reverse the transformation and replace the image block in the image.

Watermark detection is easily performed by carrying out the above operations on the original image and the watermarked image in parallel and comparing the values of the coefficients. Note that the block-mean approach is a special case of the above. If the Discrete Cosine Transform (or "DCT") is used in step 4 above to transform the image sub-blocks then the mean value will be one of the coefficients present, although it will never be marked unless step 2 is removed.

Zhao and Koch [12] have 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). Zhao and Koch also take the precaution of placing the blocks at random positions in the image in order to make a successful attack by an enemy less likely.

3.1.2 The number of bits The first stage in embedding a bit stream in an image is to determine the number of bits that can be placed into a given image block. A very simple method based on Parseval's Theorem [2] will now be described.

In a highly textured image block energy tends to be more evenly distributed amongst the different DCT coefficients. In a flat featureless portion of the image the dominant energy components tend to lie at the low frequency end of the spectrum.

As stated above, the aim is to place more information bits where they are least noticeable. This may be accomplished by using a simple thresholding technique. The steps are as follows:

1. Sort the DCT coefficients in order according to absolute magnitude.

Figure 1: *Standard grey scale image of Lena. The size of image is 512 x 512 pixels.*

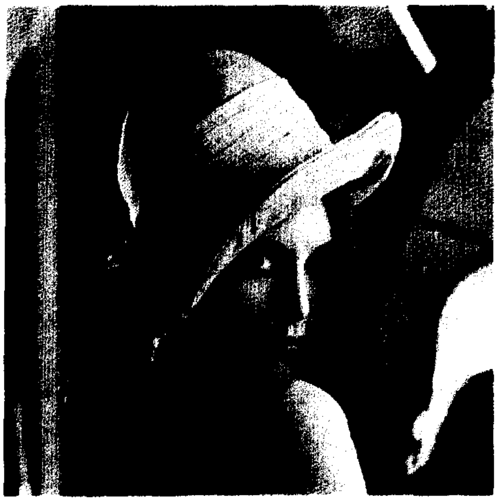

Figure 2: *Lena weakly watermarked using blocks with borders.*

2. Starting with the largest, sum the energies in each component, until a predetermined threshold (usually a simple proportion ϵ of the total energy) is exceeded.

3. Set the number of bits to be placed in this block equal to the number of components required to exceed the threshold.

This approach of placing bits where they are least visible can be a potential weakness. Lossy image compression algorithms are designed to disregard redundant information. Information bits placed within textured areas of the image are therefore more vulnerable to attack. Therefore, there is a compromise to be reached between hiding a large number of information bits where they can least be seen, but where they can be attacked by image compression algorithms, or placing fewer bits on less textured but safer portions of the image. This may be achieved by opting for a moderately low value of threshold (*e.g.* $\epsilon \approx 0.7$).

It is worth noting that the number of bits that can be encoded using image transforms far exceeds that of the block-mean approach. The expected capacity is in the order of 1000 bits for a typical image. In the case of Zhao and Koch's method 8 bits of information are encoded into each 8 × 8 block. If the blocks are tiled over the image then overall one could obtain a maximum code rate of 0.125 bits/pixel.

3.2 OTHER TRANSFORMS

The DCT is not the only image transform that may be used for watermarking. Other transforms that may be used include:

Walsh transforms: The Walsh-Hadamard transform [8, 13] can be viewed as a generalization of the block-mean approach described above. In this composed entirely of elements with value 1 or −1 only. The Walsh-Hadamard transform can be implemented using as a fast algorithm.

Wavelet transforms: The wavelet transform has been shown to give good compact representation of image texture. This suggests that it may have powerful watermarking properties. Fast wavelet transforms exist and are described and implemented by Press *et al.* [10].

The FFT: The FFT may also be applied. The great advantage of the FFT is that it allows the separation of magnitude and phase for modulation purposes.

In each of these transformations it is assumed that the block size is an integer power of two.

It is important to note the differences between the aims in image compression and in designing watermark transmission pulses. In image compression one is given a number (hopefully a small number) of coefficients with which to reproduce a good approximation to the original image. A small change in the coefficients should make little difference to the approximation to the image. However, the reverse does not necessarily hold since a small change to the image can result in a large change in the coefficients. This kind of behaviour is obviously extremely undesirable when the embedded information depends on the value of these coefficients. The severity of this effect depends on the image transforms being used. Ill-conditioning tends to be much more severe for image transformations whose basis images are data-dependent (e.g. the singular value decomposition or SVD). Image transformations with fixed basis images (e.g. DCT and wavelet transforms) tend to

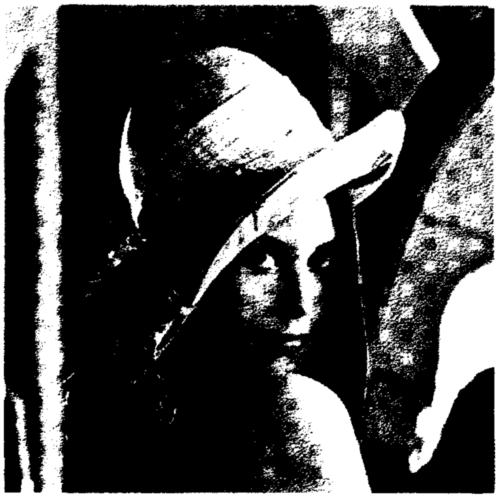

Figure 3: *Lena strongly watermarked using blocks with borders.*

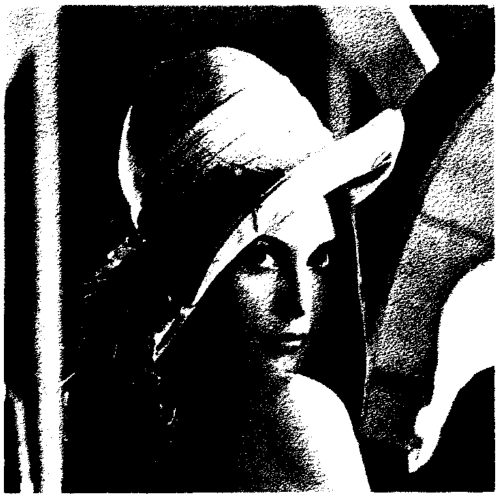

Figure 4: *Lena watermarked using fourth order Daubechy wavelets.*

3.3 OTHER ISSUES

The material in this paper thus far has described methods that may be used for placing a watermark in an image. However, we have not addressed other components in the watermarking problem, namely the reliable and secure transmission of the watermark.

Reliable communication was proven by Shannon [14] to be theoretically possible providing the information rate does not exceed a threshold known as the channel capacity. The Shannon limit may be approached by applying error control codes. Error control coding and modulation although often treated separately are in fact closely related. For example, in the implementation of the watermarking algorithm described above, the process of using only selected coefficients and ignoring others, is an example of a spherical code. In a spherical code [11] the points in signal space lie on the surface of a sphere whose radius is determined by the energy content. This code has mild error correcting properties with the result that low values of energy threshold result in significantly improved performance in the transmission of the mark. More robust error correction techniques can be employed if necessary. Methods for error control coding are described by Sweeney [15], Chambers [11] and Blahut [16].

In addition to reliability there may also be a need for security. Many different encryption algorithms exist to carry this out. A good introduction is presented by Chambers [11]. A more mathematical treatment of the subject is given by Konheim [17].

4 RESULTS

Figure 1 shows a standard image without a watermark. Figure 2 shows the same image watermarked using bidirectional coding and the blocks with borders method described above. The inner block size is 12 and the depth of modulation is 3. Figure 3 shows the same image strongly

The mark is for all intents and purposes invisible in figure 2 but may be detected quite readily even after lossy compression and scanning has been carried out. The watermark conveys 441 bits of information and the standard message reads: "012345 This is a watermark...".

Figure 4 shows "Lena" watermarked using the Daubechy Wavelet Transform. The block size is 8 and the depth of modulation is 5. The watermark conveys 12882 bits of information. The standard message is repeated to occupy all the available capacity. Note that the presence of the mark introduces no visible degradation.

The question arises as to what a watermark actually looks like. Figure 5 shows the difference between the wavelet marked version of the standard image and the original, scaled by a factor of 32. Figure 6 shows the image of a watermark produced using the DCT. As in the case of the wavelet watermark, the block size is 8 and the depth of modulation is 5. The number of bits encoded in the DCT watermark equals 9342 and the standard test message is repeatedly encoded as before.

The DCT watermarked image was compressed using JPEG with default settings. The quality factor was set to 90 and no smoothing was used. The compression ratio was 14:1. The binary bit pattern in the watermark was recovered with a bit error rate of 14%. Since the errors tended to occur in bursts it was actually possible to decipher ASCII characters from the raw bit stream without resorting to error-control codes. The same experiment was repeated using images marked using Daubechy wavelets and the Walsh-Hadamard transform. The corresponding bit error rates were 18.5% and 20.5% respectively.

It is apparent upon examining the watermarks in figure 5 and figure 6 that the transform based marking schemes possess a number of desirable features. First, one can mark according to the distribution of energy within the coefficients. In this way, one can place watermarks where

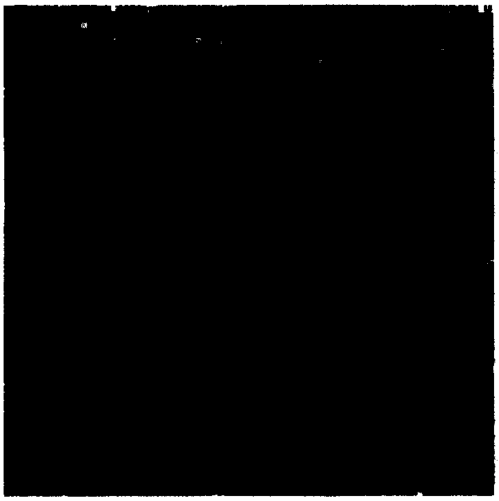

Figure 5: *The watermark produced using Daubechy wavelets.*

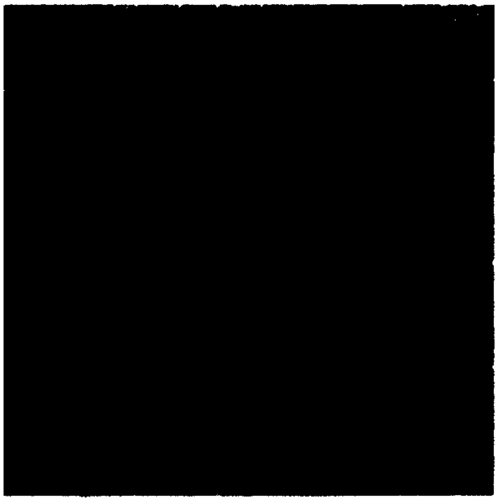

Figure 6: *The watermark produced using the Discrete Cosine Transform.*

As a result, the watermark exhibits a ghost-like resemblance to the original image. It is also very interesting to note that the watermark pattern on the flat regions of the image (such as Lena's shoulder) bears a superficial resemblance to military camouflage. Second the watermark is irregularly distributed over the entire image sub-block which makes it more difficult to detect and for enemies in possession of independent copies of the image to decode and read the mark.

5 CONCLUSION

This paper has outlined a scheme for embedding robust watermarks onto digital images. The watermarks are designed to be invisible even to a careful observer but contain sufficient information to uniquely identify both the origin and intended recipient of an image with a very low probability of error.

Future work will involve the further development of robust error correction codes and digital signature techniques. In addition, the authors will attempt to envisage possible attacks on the integrity and security of the mark and to devise suitable countermeasures.

References

[1] Brassil, J., Low, S., Maxemchuk, N., and O'Gorman, L., 1994, "Electronic Marking and Identification Techniques to Discourage Document Copying," in "INFOCOM 94".

[2] Bissell, C. C. and Chapman, D. A., 1992, "Digital Signal Transmission". Cambridge University Press.

[3] Kurak, C. and McHugh, J., 1992, "A cautionary note on image downgrading," in "Proc. 8th Annual Computer Security Applications Conference", (San Antonio).

[4] Walton, S., 1995, "Image Authentication for a Slip-

[5] Dautzenberg, C. and Boland, F. M., "Watermarking Images," tech. rep., Department of Electronic and Electrical Engineering, Trinity College Dublin, 1994.

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

[7] Caronni, G., "Assuring Ownership Rights for Digital Images." Submitted for publication in ASIACRYPT'94, 1994.

[8] Clarke, R. J., 1985, "Transform Coding of Images". London: Academic Press.

[9] Rao, K. R. and Yip, P., 1990, "The Discrete Cosine Transform: algorithms, advantages, applications". Academic Press.

[10] Press, W., Teukolsky, S., Vetterling, W., and Flannery, B., 1992, "Numerical Recipes in C". Cambridge University Press, second ed.

[11] Chambers, W. G., 1985, "Basics of Communications and Coding". Oxford Science Publications, Clarendon Press Oxford.

[12] Zhao, J. and Koch, E., "Embedding Robust Labels Into Images For Copyright Protection," tech. rep., Fraunhofer Institute for Computer Graphics, Darmstadt, Germany, 1994.

[13] Pearson, D., 1991, "Image Processing". Essex Series in Telecommunication and Information Systems, McGraw-Hill.

[14] Shannon, C. E., 1948, "A mathematical theory of communication," Bell Sys. Tech. J., 27, 379–423 and 623–56.

[15] Sweeney, P., 1991, "Error Control Coding: An Introduction". Prentice-Hall.

[16] Blahut, R. E., 1983, "The theory and practice of error control codes". Addison-Wesley.

[17] Konheim, A. G., 1981, "Cryptography: A primer". New York: John Wiley and Sons.