Robust Public Key Watermarking of Digital Images

Balas Natarajan
Computer Systems Laboratory
HPL-97-118
October, 1997

E-mail: balas@hpl.hp.com

robust,
watermarking,
invisible

This paper addresses the problem of watermarking a digital image, in order to detect or verify ownership. An invisible and resilient watermark is one that is not ordinarily visible and is robust in the face of common image manipulations such as lossy compression, scaling, cropping, rotation, reflection, and brightness/contrast adjustment. A watermarking scheme is non-invertible if a valid watermark cannot be subtracted from an image, within the framework of the scheme. We propose an invisible, resilient and non-invertible watermarking scheme that uses public key signatures, and is computationally inexpensive to create, detect and verify.

Internal Accession Date Only

2 Introduction

There are several ways of classifying watermarks for images. One classifications partitions them into visibile and invisible watermark. Another classification partitions them into resilient watermarks and integrity watermarks. Visible watermarks are created by simply blending the watermark image with the image to be protected. The watermark signal is typically a logo or copyright information. Visible watermarks are simple to create, detect and verify, but are of course obtrusive. Invisible watermarks are unobtrusive additions to an image whose presence can be checked by a verification algorithm. Their unobtrusiveness makes them more desirable, but also more difficult to create and verify. Orthogonal to the visible/invisible classification, watermarks can be classified as resilient or integrity. A resilient watermark is one that is resistant to tampering via operations such as scaling, cropping and smoothig the image. In contrast, an integrity watermark is like a "checksum" and detects changes to the image such as cropping scaling or tampering of the pixel values. Wong [4] reported recent work on a block-based integrity watermking scheme. In this paper, we are concerned with resilient, invisible watermarks only.

Previously, Cox et al., [1], proposed a watermarking scheme for images. In their scheme, a two-dimensional spread spectrum signal is added to the image. The spread spectrum signal is created by using the binary key of the owner of the image to modulate the spectral coefficients of the watermark signal. The binary key would typically consist of 128 bits, and the low-order 128 coefficients of the watermark signal would be derived as modulations of these bits. The high order frequencies of the watermark signal would be zero. Inverting this modulated spectrum yields the watermark signal, which is then added to the original image to create the watermarked image. To detect the watermark in a given image, the original image is subtracted from the given image, and the correlation of the difference image and the watermark signal is computed. If the correlation is high, the given image is deemed to be watermarked.

The scheme of Cox et al. suffers from the following limitations (1) The original image is necessary to detect the presence of the watermark. This places a substantial storage or transmission burden, when a large number of sizeable images are involved. (2) Operations such as cropping and scaling pose a considerable computational burden on the verification process, since the image to be tested has to be registered exactly against the original. Furthermore, the scheme treats cropped images by embedding them in a full-sized image before detection, leading to considerable obfuscation in the detection process. (3) The watermarking scheme is invertible, leading to ownership disputes as described in [3]. In particular, it possible for a malignant party to subtract his watermark from a watermarked image, and declare the resulting image to be his original. (4) Cox et al. do not offer a mechanism for creating and managing watermarks. If the same watermark is used to protect several images, compromise in the secrecy of the watermark will compromise all the images. (5) Changes in brightness/contrast will fool the verification algorithm easily, owing to the subtraction against the original image.

The goal of our work is a watermarking scheme that combines the strength of the spread spectral technique introduced by [1], while mitigating its weaknesses as identified above. To this end, we propose a spread spectral scheme that uses public-key signatures. The

scheme uses watermarking bands that are strategically placed at important locations in the image, so that any reasonable portion of the image will contain one or more of the bands. Watermark bands are computationally less expensive to detect than two-dimensional full-image watermarks, particularly in the face of cropping and scaling operations. Furthermore, our scheme does not require the original image in order to detect the watermark, only the watermark signal (or key) and a corresponding reference vector are required. The use of public-key signatures to create the watermark allows two important features of the scheme. Firstly, the watermark is not invertible, eliminating a weakness of the scheme proposed by Cox et al. Secondly, it allows the owner of an image to prove ownership to an impartial judge, by submitting to the judge the image in question, along with the original image, the watermark vector, and his public key. The owner of the image does not reveal his private key, and therefore does not compromise his ability to protect other images that he might have watermarked.

3 Proposed Method

Before we give the details of our watermarking scheme, we briefly review public key signatures, [2]. There are several public key signature algorithms, the most common being RSA, and DSS. Each user possesses a pair of keys, a private key and a public key. The private key is kept secret and is known only to the user. The public key can be distributed widely. The key pair has the property that a message encrypted with either key can only be decrypted with the other key. This property can be used to sign messages. If a user encrypts a message with his private key, then it can only be decrypted with his public key. Since the private key is known only to the user, it is established that he and only he, could have encrypted the message, i.e., the user has 'signed' the message. Since the strength of the signature is dependent on knowing that the public key of the user is genuine, public keys are notarized or certified by third party notaries. Also, it suffices for the user to encrypt a hash of the message to create a signature, rather than encrypt the entire message, in the interest of computational efficiency. The hash is called the 'message digest' and should be computed using a strong hash function, such as the 128-bit MD5.

We can now give the details of our watermarking scheme. We describe each of the operations of insertion, detection and verification in succession. In brief, the owner of an image adds a watermark using the insertion procedure. The detection procedure allows for the inserted watermark to be detected by any party so authorized by the owner. The verification procedure allows the owner to prove ownership to an impartial judge.

3.1 Watermark Insertion Procedure

To watermark a given original image I, of M rows and N columns:

(1) Message Digest Step: Compute a message digest of the bits in I, using a standard message digest algorithm such as MD5, [2].

(2) Encryption Step: Encrypt the message digest with the private key to create the owner's public key signature S.

(3) Modulation Step: Use the bits of the signature to construct a watermark signal as follows. Construct the vector U of N entries, where entries 2 through 65 take values of of -1 or +1, depending on whether the corresponding bits of the signature S are 0 or 1. The remaining entries of U, i.e., entries 1 and 66 through N are all zero. The vector U corresponds to the spectral coefficients of the watermark vector, in order of increasing frequency. Compute the fourier transform of U, to obtain the watermark vector V.

(4) Orthogonalization Step: Select b, say $b = 16$, contiguous rows of the image I at random. Average these rows to construct an average row vector A. A is called the reference vector. Orthogonalize V with respect to A to obtain the watermark vector W,

$$W = V - (V.\hat{A})\hat{A} \ , \tag{1}$$

where $\hat{A}$ is the unit vector along A.

(5) Watermarking Step: Add a small scaled version of W back to each of the b rows selected, to obtain the watermarked image. The scaling factor is modulated as shown below across the b rows, to ensure that the strength of the watermark varies smoothly. Let I_1, I_2....I_b be the b rows of the image that are selected. Viewing these as row vectors of pixels, set

$$I_i = c \cos\left(\frac{2\pi i}{b}\right) W \ , \tag{2}$$

where c is a small constant that controls the strength of the watermark. Typically, c is chosen so that the watermark signal is roughly -40dBPSNR. Additional watermarks can be added at other locations of the image as desired, by repeating the orthogonalization and watermarking steps using the same key vector V. Store V and each of the reference vectors A so created.

3.2 Watermark Detection Procedure

In this section we present our procedure for detecting the presence of a watermark in a given image J. We require the watermark vector V and the reference vector A. We do not require the original image I.

(1) Correlation Step: Orthogonalize V with respect to A, to obtain vector W. Successively examine each block of b rows in the image J. Average the rows to get an average vector B. Orthogonalize B with respect to A, to obtain vector X. Compute the correlation between W and X. Take the maximum of the correlation over all contiguous blocks of b rows.

(2) Search Step: Search the space of cropping and scaling factors, and repeat the Band Search Step for each point in the space. Also, repeat for lateral reflections of the image around the vertical axis. Compute the maximum correlation obtained over all of the above computations.

(3) Reference Correlation Step: Synthesize, say, one hundred random candidate watermark vectors, with the same spectral properties as V. Compute the correlation of each of these candidate vectors with the image J, at the location, scale, and crop factors at which the correlation of V was maximised.

(4) Decision Step: Compare the correlation obtained with V, against the correlation obtained for these random vectors. If the former and the latter are far apart, then it is very likely the the image J contains the watermark V, as opposed to one of the randomly synthesized watermarks. If not, we will deem that the image J does not contain the watermark V.

3.3 Watermark Verification Procedure

In this section we describe our procedure for establishing the presence of a watermark with a neutral judge. A person claiming ownership of an image would present to the judge the watermarked image J, the corresponding original image I, the signature S of the hash of the image, and his public key certificate. The claimant would also declare the location in the image J where the watermark is claimed, and the scaling and cropping factors of J with respect to the original I. The judge would carry out the following steps.

(1) Signature Verification Step: Compute a message digest of the bits in I. Decrypt S with the public key presented. The two bit strings obtained above should match. If not, reject the ownership claim.

(2) Watermark Detection Step: Construct the watermark vector V from the signature S. At the specified location of the image J, compute the correlation of the watermark vector after compensating for the scaling and cropping factors specified. Use the original image I for computing the corresponding reference vector.

(3) Reference Correlation Step: Synthesize, say, one hundred random candidate watermark vectors, with the same spectral properties as V. Compute the correlation of each of these candidate vectors with the image J, at the location, scale, and crop factors specified.

(4) Decision Step: Compare the correlation obtained with V, against the correlation obtained for these random vectors. If the former and the latter are far apart, then it is very likely the the image J contains the watermark V, as opposed to one of the randomly synthesized watermarks. If not, deem that the image J does not contain the watermark V.

4 Experimental Results

The image shown in Figure 1 is the original image of size 384x256 pixels, and that of Figure 2 is the same image after a single watermark was inserted. In creating the watermark, the Message Digest Step and the Encryption Step of the watermarking procedure were skipped. Instead, we simply generated 64 random bits and used them to modulate the spectral coefficients in the Modulation Step. The strength of the watermark is -45dBPSNR over the rows along which it was inserted. Figure 3 is the difference between the watermarked image and the original image and gives the location and nature of the watermark signal. To enhance the clarity of the figure, the watermark signal has been amplified 4 times in amplitude.

In Figure 4, we show the correlation obtained by the detection algorithm for the watermarked image of Figure 2. The spike in the middle of the plot is the correlation of the true key, while the rest of the plot is the correlation obtained by randomly generated keys of the same spectral bandwidth.

Figure 1: Original image.

Figure 2: Watermarked image.

Figure 3: The watermark signal. (Normalized and amplified 4x).

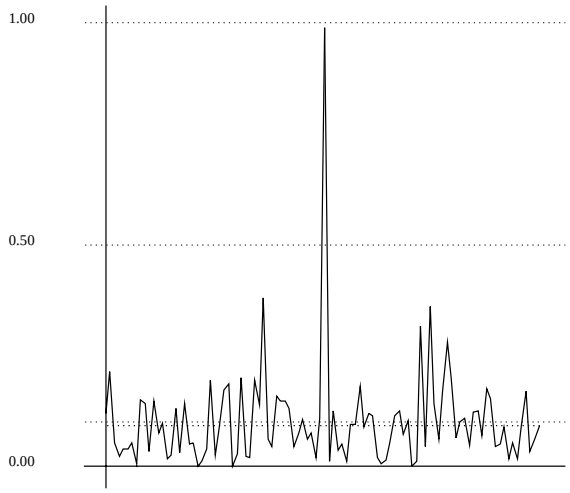

Figure 4: Correlation spread of Figure 2. Central spike corresponds to true watermark, and the other points to randomly generated candidate watermarks.

Figure 5: Cropped and JPEG compressed watermarked image.

To test the robustness of the watermark, we cropped the watermarked image to 274x176 pixels, Figure 2, JPEG compressed it with substantial loss achieving a compression ratio of 28.6, and then decompressed it to obtain Figure 5. The plot of Figure 6 is the correlation spread obtained by the detection procedere for the image of Figure 2. Notice that the spike is not as strong as that of Figure 4, but is clearly substantial despite the cropping and JPEG loss.

Figure 7 (241x153 pixels) is a downscaled version of the watermarked image Figure 2 (384x256 pixels). Figure 8 is the correlation spread obtained by the detection procedere for the image of Figure 7. Notice that the watermark is resilient to the scaling process.

As another test of robustness, we set to zero the five least significant bits of each eight-bit pixel in the watermarked image, to get the truncated image Figure 9. Figure 10 is the correlation spread obtained by the detection procedure for the image of Figure 7. Notice that the watermark is surprisingly resilient to the truncation process. The resilience stems from the fact that the spread spectral watermark is a redundant encoding of the watermark

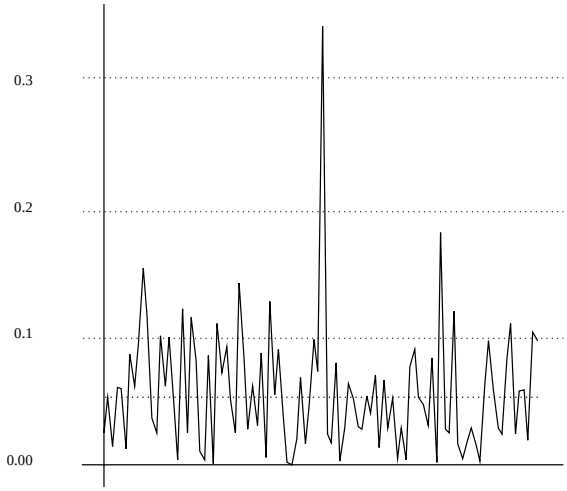

Figure 6: Correlation spread of watermarked image after cropping and JPEG compression. Central spike corresponds to true watermark, and the other points to randomly generated candidate watermarks.

Figure 7: Downscaled version (241x153 pixels), of the watermarked image (384x256).

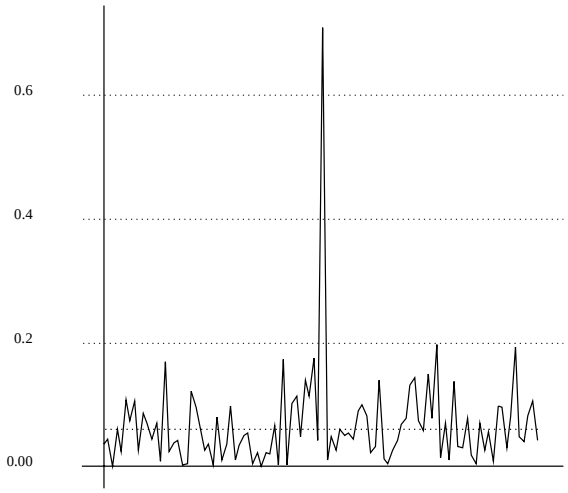

Figure 8: Correlation spread of scaled watermarked image. Central spike corresponds to true watermark, and the other points to randomly generated candidate watermarks.

information, and that the addition of the watermark changes the third most significant bit of sufficiently many pixels to enable the watermark to survive the truncation process.

5 Conclusion

We presented an invisible, resilient and non-invertible watermarking scheme that uses public key signatures. The watermark is computationally inexpensive to create, detect and verify.

6 Acknowledgements

Thanks to V. Bhaskaran, C. Herley, J. Hogan and M. Khansari for the insightful discussions.

References

[1] I.J. Cox, J. Kilian, T. Leighton, and T. Shamoon, "Secure Spread Spectrum Watermarking for Multimedia," *NEC Research*, TR-95-10.

[2] Schneier,B., *Applied Cryptography*, John Wiley and Sons, 1993.

[3] Craver, S., Memon, N, Yeo, B-L, and Yeung, M., "Can Invisible Watermarks resolve rightful ownerships?" *IBM Research Report*, RC20509, July 1996.

[4] Wong, Ping Wah, "Image Security and Watermarking" Imaging Tech. Dept., HP Labs, October 1996.

Figure 9: Watermarked image with five least significant bits truncated in each eight-bit pixel.

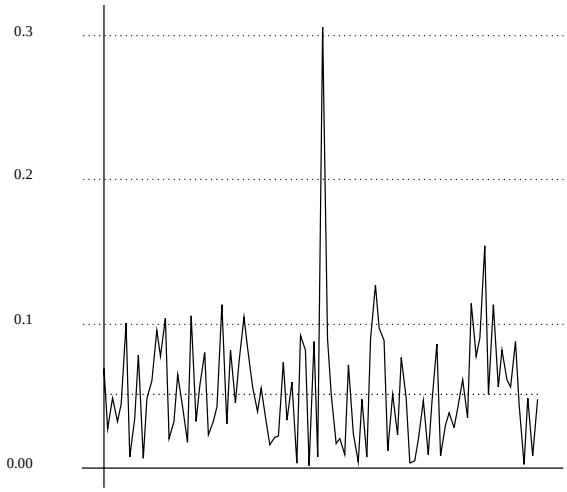

Figure 10: Correlation spread of watermarked image after truncation of the five least significant bits in each eight-bit pixel. Central spike corresponds to true watermark, and the other points to randomly generated candidate watermarks.