PICS 1999

Robust watermarking using argument modulation

Poorvi L. Vora
Hewlett-Packard Laboratories
1501 Page Mill Road
Palo Alto, CA 94304. 650-857-2457
poorvi@hpl.hp.com

Abstract

We describe the use of argument modulation for the robust watermarking of multimedia documents. Unlike amplitude modulation, argument modulation uses position in a (transformed) sequence of multimedia bytes to encode information. The distinctive features of the scheme are: it requires the original image for detection; the watermark is invisible without requiring the absence of correlation in the watermark (so, for example, the watermark can be a logo); it is consistent with private key encryption; it is non-invertible; it is not incorporated additively. The last-mentioned feature makes this watermark more difficult to remove than other watermarks. We show with examples that argument modulation is exceptionally robust to common accidental attacks and we describe the difficulties of using intentional attacks to destroy the watermark.

1. Introduction

Recent increase in the accessibility of computational power, magnetic storage and the internet has led to a profusion of digital multimedia documents and their distribution over the internet. With the ease of making perfect copies of digital multimedia documents and distributing them, it is necessary to provide some kind of protection for copyright owners. Robust, invisible watermarking of multimedia documents is being suggested as a means of such protection. It has other uses as well [1]. In this paper, we describe a robust watermarking method in detail for still images and the extension to sound, video and any other multimedia document is straightforward.

For robustness the watermark must be inserted in a region of perceptual significance (low and middle frequencies) [2, 3, 4, 5] so that any damage to the watermark includes damage to the perceptual quality of the image. Further, for the watermark to be invisible, it must not be strong in regions of great perceptual significance (low frequency regions). This implies a clear trade-off between watermark perceptibility and robustness. There is considerable literature on addressing this trade-off for methods that are additive in the watermark. There are also trade-offs among robustness, visibility and information content of a watermark, though there is not much literature addressing this.

Most published methods for robust, invisible, watermarking [2, 4, 5, 6, 7] add the watermark to transform or pixel values, and these methods are hence linear with respect to the watermark. Much is known about linear systems and this knowledge is available to an adversary, making most existing methods more vulnerable to attempts at removal [8]. Further, additive methods are particularly vulnerable to additive noise and to changes in brightness. Lastly, additive methods increase the image power, even if only slightly.

Other methods [9, 10] enforce constraints on the ordering or differences or linear combinations of transform (block Discrete Cosine Transform - DCT, usually) values. While these methods are resistant to noise, they are not particularly robust to other operations. One other method [3] hides the information in the phase of the Discrete Fourier Transform (DFT). Finding the phase is a very messy process, and the computational complexity and numerical error of such a method are undesireable. Further, the phase is extremely susceptible to translations of the original image. IBM's published, copyright protection, invisible watermarking method [11] consists of multiplying the amplitude of the image by the watermark. This method, like all the other methods, is vulnerable to attack because it encodes information in the amplitude of the image.

While the details of the different methods vary as each author proposes different ideas for invisibility or robustness, and ways to obtain some degree of both, the crucial similarity among all these methods except for [3] is that they are based on hiding the information using the amplitude of the pixel values or some linear transformation of them.

We propose a method completely different from the exist-

ing methods. We propose that instead of hiding information in the *amplitude* of a linear transformation, it be hidden in the *argument* of an invertible linear (or non-linear) transformation of the image. This makes the watermarking technique a non-linear operation with respect to the watermark. We expect this method to be more robust to both enhancement of the image (or accidental attacks) and (intentional) attacks. We also expect this method to hold more information for the same degree of watermark perceptibility.

While benefitting from the advantages of being different from the existing schemes, we can use all the past work done in identifying frequency ranges of perceptual importance [2, 5] as well as work done on extending watermark detection on perfectly registered images to watermark detection on cropped, rotated and/or scaled images [5] - because this work is not specific to amplitude watermarking.

The method described here is robust to common image transformations like compression, contrast enhancement, brightness variation, conversion to grey-scale, blurring and sharpening as long as these operations preserve the perceptual quality of the image. This method has sufficient in-built protections to make conscious watermark removal by an adversary a difficult task. The watermark is non-invertible [12] - i.e. an adversary cannot undo it and claim that the original image with the watermark subtracted is another valid image within the framework of the scheme. Further, unlike other watermarking schemes, considerable correlation in the watermark does not compromise the invisibility of the watermark. At the same time, encoding the correlated message in an uncorrelated manner should only strengthen the security of the scheme.

The proposed method provides the following advantages over most other schemes:

1. It is possible for the watermark to have considerable structure and still be imperceptible because it is not added to the amplitude of any linear tranformation of the image.
2. Watermark removal is more complicated because it involves going to the argument in the transform domain.
3. It is possible that the manner of incorporation also allows incorporation of more information in the watermark for a given degree of watermark robustness and perceptibility, but this is not yet clear.

This paper is organized as follows. Sections 2 and 3 describe the method in outline and in detail respectively. Section 4 describes the robustness of the method by detailing the results of transformations of an image watermarked using the method. Section 5 presents conclusions.

2. Basic Method

Let f(i,j) represent the $(i,j)^{th}$ pixel in the image. Let F(k,l) represent the $(k,l)^{th}$ value of the transformed image f(i,j). For example, F could be the DFT or block-DCT value of the image. It could also be the Walsh, Hadamard, Discrete Hartley or Discrete Wavelet transform value of the image. It could also be any other, hypothetical, invertible (even if non-linear) transform of the image. Let g(i,j) be the watermarked image and G(k,l) its transformed image. Let w(i,j) be the watermark.

2.1 Insertion

Choose a range of argument values in the transform domain that retain sufficient perceptual information so that destroying the watermark will destroy image quality, but make sure that these are not values that are perceptually dominant [7]. From among these values, choose a proper subset based on the private encryption key. Let $\mathcal{S}$ denote this subset.

$$G(k,l) = \quad (1)$$

$$\begin{cases} F(k+\alpha \times w(k,l) - \gamma, l+\beta \times w(k,l) - \delta) & (k,l) \in \mathcal{S} \\ F(k,l) & \text{else} \end{cases}$$

where α, β, γ and δ provide affine conversions between values of w(k,l) and values of the argument shift. Obviously, the conversions do not have to be affine. Further, the watermarks for the k and l directions could be distinct, reducing redundancy and robustness but increasing hidden information content. If w(k,l) is a colour watermark, each frame of the watermark can be inserted in each frame of the image in the above manner. If the transform of a single image consists of two different images (as in the DFT) the same watermark or different parts of the watermark can be inserted in each transformed image with well-known tradeoffs among redundancy, information content and robustness.

To reduce the possibility of an attack where an adversary claims the watermarked image as an original and the negative watermark as their watermark (and hence the original image as their watermarked image) [12] we change equation (1) to the following:

$$G(k,l) = F(k + \Delta k, l + \Delta l) \quad (2)$$

where:

$$\Delta k = +/- (\alpha \times w(k,l) + \gamma) \quad (3)$$

depending in some manner on F(k,l) - for example, on the value of the i^{th} bit of F(k,l). Similarly with Δl. Obviously, there are other ways of ensuring that the watermark insertion into the argument is non-invertible (and image-dependent).

2.2 Detection

Detection of the watermark is performed using the original image and the range of argument values in which the watermark was inserted. The transformed candidate image I(k,l) over the specified range of arguments is compared to the values of G(k,l) from equations (2) and (3) for every possible value of w(i,j). If w(i,j) takes on values from the set $\{S_k\}_{i=1}^{i=N}$, then its estimate, $\hat{w}(i,j)$ is:

$$\hat{w}(i,j) = \tag{4}$$

$$\underset{x \in S_k}{min} |G(i,j) - F(i + \alpha \times x - \gamma, j + \beta \times x - \delta)|$$

This may be calculated separately for each frame in which the watermark is inserted. If the watermark insertion has some degree of redundancy - for example a binary watermark inserted in the DFT of a colour image will be inserted in six frames - the watermarks detected from each image may be added together and the resulting image thresholded, or the detected watermarks may be logically anded together. A redundancy helps make the method far more robust. The redundancy can be used in a far more stringent form by looking at all six frames together and choosing a best fit for all six frames simultaneously. Different ways of using the redundancy do not change the method in any fundamental manner, and error control coding to encode the watermark would be extremely useful and would fit into the framework we have described.

3. Details

1. Generate watermark.

   The watermark is a two-dimensional distribution of intensities. It may be colour, grey-scale or binary with increasing robustness. We have found binary to be the most practical. The watermark may be an encrypted, error-control coded message or an unencrypted, unencoded company logo. The former is more secure. The size of the watermark and whether it is binary, grey-scale or colour depends on the required degree of robustness, the required amount of information to be hidden, and the computational complexity of the detection procedure which depends on the number of levels in the watermark.

   The intensity distribution in space of the watermark (i.e. the watermark image) affects the robustness of the technique - especially if the watermark is highly correlated with itself. Hence, the values and their positioning with respect to perceptually significant transform arguments affects the perceptibility and robustness of the watermark. Thus watermark design depends on both the nature of the information to be transmitted and on the optimal watermark for the expected data set of images.

   For a 512 × 512 image we used a binary watermark of size 150 bits.

2. Insert watermark.

   - Choose the transform in which to insert the watermark. The results we present here use the DFT. Other possibilities include DCT and block-DCT, wavelet-based transforms, Walsh, Hadamard and Discrete Hartley transforms. As with other watermarking methods, global transforms provide more robustness and also more perceptibility of the watermark.
   - Choose the argument range for embedding the watermark. This choice can be standard for all images or can be image-dependent. We are working on a method for an image-dependent choice. For the DFT this choice will focus on the 'middle range' of frequencies because of the trade-off between robustness and perceptibility. We used the frequency range around 0.2 times the bandwidth.
   - Choose the step-size in argument that corresponds to a unit value of the watermark image, and also choose what zero corresponds to - i.e., choose α, β, γ and δ. In the images we show here we used a step size of 2 units corresponding to a white, and 0 units for a black value. The step size too can be image-dependent and can also change over a given image depending on the perceptual importance of a certain range of arguments. Both step-size and frequency range should be chosen so that the watermark is just imperceptible and so that detection is close to unique. We are working on details of this aspect.
   - Insert watermark as described in equation(2) by first finding the transform, then inserting the watermark in the argument and then finding the inverse transform.

3. Detect watermark

   Using the original image, the argument range for the original image and the argument step size of the original image, retrieve a detected watermark from the candidate image as described in equation (4). Compare this to the inserted watermark. Do this for many different estimates of cropping, scaling, rotation and translation as in [5].

4. Results

Here we present the results obtained on one sample image using the parameters specified in the previous section. In each of the real and imaginary frames in each of the colour frames, we inserted three copies of the watermark. We performed common image processing operations on the watermarked image to test the robustness of the watermark. After each operation, we detected the watermark by detecting three marks in the magnitude frame of the DFT of each colour frame and then finding the average of the nine watermarks. As a performance measure, we counted the fraction of bits that were recovered correctly. Table 1 shows the results.

Detecting the watermark after geometric transformations requires an estimate of the transformation. Some methods [5] require the estimate be performed by a human. We padded the smaller images using zeroes after cropping. The scaling results shown here assume perfect registration and are resized to the original using subsampling or pixel replication depending on whether the watermarked image size is more or less than that of the original.

The method is least robust to geometric transformations - this is true of almost all robust watermarking methods. The method is remarkably robust to all other transformations.

5. Conclusions and Future Directions

Hiding the information in the argument of an invertible transformation of the image appears to be robust to most transformations. The method described here can be improved upon by developing more complex image-dependent techniques to determine the parameters (step-size and whether it is constant or changes over the image; range of arguments to be changed; binary, grey-scale or colour watermark; type of watermark) of the method. Further, the method can easily be applied to audio and video streams, where application scenarios for watermarking as a means of copyright protection are more plausible. Lastly, a framework for the study of the robustness of, and information content in, non-linear watermarking schemes is necessary before one can theoretically compare the robustness of this scheme with that of schemes that modify amplitude.

References

[1] Nasir Memon and Ping Wah Wong, 'Protecting Digital Media Content', Communications of the ACM, vol. 41, no. 7, July 1998.

[2] F.M. Boland, J.J.K. Ó Ruanaidh and C. Dautzenberg, 'Watermarking Digital Images for Copyright Protection' IEE Int. Conf. on Image Processing and Its Applications, pp 321-326, Edinburgh, July 1995.

[3] J.J.K. Ó Ruanaidh, W.J. Dowling and F.M. Boland, 'Phase Watermarking of Digital Images' IEEE Int. Conf. Image Processing, Vol III pp 239-241, Lausanne, Switzerland, Sept 1996.

[4] Bo Tao & Bradley Dickinson, 'Adaptive Watermarking in the DCT Domain', IEEE Conference on ASSP, April 1997.

[5] Geoffrey Rhoads, 'Steganography Methods Employing Embedded Calibration Data', U. S. Patent 5,636,292, Jun. 3, 1997.

[6] I.Pitas and T.H. Kaskalis, 'Applying Signatures on Digital Images', IEEE Workshop on Nonlinear Image and Signal Processing, Neos Marmaras, Greece, pp. 460-463, June 1995

[7] I.J. Cox, J. Kilian, T. Leighton and T. Shamoon, 'A Secure, Robust Watermark for Multimedia', Workshop on Information Hiding, Newton Institute, Univ. of Cambridge, May 1996.

[8] Ton Kalker, Jean-Paul Linnartz and Marten van Dijk, 'Watermark estimation through detector analysis', IEEE Int. Conf. on Image Proc., Chicago, 1998.

[9] E. Koch & J. Zhao, 'Towards Robust and Hidden Image Copyright Labeling', Proc. of 1995 IEEE Workshop on Nonlinear Signal and Image Processing (Neos Marmaras, Halkidiki, Greece, June 20-22, 1995), pp. 452-455.

[10] A. Bors and I. Pitas, 'Image Watermarking Using DCT Domain Constraints', 1996 IEEE International Conference on Image Processing (ICIP'96), Lausanne , Switzerland, vol. III, pp. 231-234, 16-19 September 1996

[11] Gordon W. Braudaway, 'Protecting Publicly-Available Images with an Invisible Watermark', IEEE Int. Conf. on Image Proc., Santa Barbara, Oct. 1997.

[12] S. Craver, N. Memon, B. L. Yeo and M. M. Yeung, 'Can Invisible Watermarks Resolve Rightful Ownerships?' , IBM Research Technical Report RC 20509, IBM CyberJournal, July 25, 1996

Table 1: Fractional Watermark Recovery

Operation Type	Operation	Strength of Degradation	Fractional Recovery
Geometric	Cropping	4%	0.71
	Rotation	5 degrees	0.61
	Translation	20 pixels	0.9
	Scaling	2	1.0
		1.1	1.0
		0.5	0.94
		0.33	0.81
Colour	Grey-scale		0.94
	Brightness Change	+50%	1.0
		+75%	0.98
		+100%	0.99
		+150%	0.97
		+200%	0.96
	Saturation Change	+50%	1.0
		+75%	1.0
		+100%	1.0
		+150%	1.0
		+200%	1.0
		+300%	0.99
		+400%	0.99
	Linear colour correction	many matrices	1.0
Spatial Domain	Dithering		1.0
	Gaussian Blurring	s.d. = 1.25 pixels	0.91
		s.d = 1.5 pixels	0.77
	Rank Order Filtering 3 by 3 window	Median (rank = 5)	0.99
		Dilation (rank = 9)	0.84
		Erosion (rank = 1)	0.87
	Unsharp mask sharpening	sharpening factor = 95%	0.99
Frequency Domain	JPEG compression	quality factor = 10%	0.89
Amplitude	Repalletization	256 colours	1.0
		16 colours	0.98
		16 colours with Floyd-Steinberg	0.91
		8 colours	0.91
		8 colours with Floyd-Steinberg	0.91
	Requantization from 8 bits/channel to	3 bits	1.0
		2 bits	0.89
		1 bit	0.79