A New Wavelet-Based Scheme for Watermarking Images

Rakesh Dugad, Krishna Ratakonda and Narendra Ahuja *
Department of Electrical and Computer Engineering
Beckman Institute, University of Illinois, Urbana, IL 61801.
{dugad,ratakond}@uiuc.edu

Abstract

A new method for digital image watermarking which does not require the original image for watermark detection is presented. Assuming that we are using a transform domain spread spectrum watermarking scheme, it is important to add the watermark in select coefficients with significant image energy in the transform domain in order to ensure non-erasability of the watermark. Previous methods, which did not use the original in the detection process, could not selectively add the watermark to the significant coefficients, since the locations of such selected coefficients can change due to image manipulations. Since watermark verification typically consists of a process of correlation which is extremely sensitive to the relative order in which the watermark coefficients are placed within the image, such changes in the location of the watermarked coefficients was unacceptable. We present a scheme which overcomes this problem of "order sensitivity". Advantages of the proposed method include (i) improved resistance to attacks on the watermark, (ii) implicit visual masking utilizing the time-frequency localization property of the wavelet transform and (iii) a robust definition for the threshold which validates the watermark. We present results comparing our method with previous techniques, which clearly validate our claims.

1 Introduction

Previous methods for watermarking images can be placed under two categories based on whether or not they use the original image during the watermark detection process. Schemes reported in [1, 2, 3] require the original image for detection, where as the method in [4] does not. Most of the reported schemes use an additive watermark, the watermark being added to the image in the spatial domain[5, 6] or in the transform domain[3, 2, 1, 4]. It is found that the transform domain watermarking schemes are typically much more robust to image manipulation as compared to the spatial domain schemes. The proposed scheme does not use the original image for watermark detection and casts the watermark in a DWT domain.

In order for a digital image watermark to be effective it should (a) be robust to common image manipulations like lossy compression, linear or non-linear filtering, scaling, cropping, collusion attacks etc. Any successful attempts at removing the watermark should noticeably degrade the image quality. (b) The watermark should be unobtrusive i.e. addition of the watermark should not degrade the perceptual quality of the original image. These are conflicting requirements. To ensure condition (a), the watermark should be added to *significant* coefficients in a suitable transform domain. But altering the significant coefficients can severely degrade the image quality thus jeopardizing condition (b).

To take care of the trade-off mentioned above, Cox *et al.* [1] suggested adding the watermark to only the *top* (largest in absolute magnitude) thousand coefficients (excluding the DC coefficient) of the DCT of the image. Since this is only a small fraction of the number of significant coefficients in a typical image there is not much perceptual degradation of the image. Detection involves retrieving the watermark by subtracting the original image from the watermarked test image and correlating the retrieved watermark with the original watermark. A sharp peak in the cross correlation coefficient indicates the presence of the watermark in the image. Using the original image in this manner can lead to complications which might prevent the watermarking scheme from ensuring rightful ownership. [7, 8] show that it is not necessary to even erase the watermark to subvert this watermarking scheme! Also no definite threshold has been defined to judge the presence of the watermark. The value of the correlation coefficient can vary significantly when the image is severely degraded, thus making it impossible to choose a universal threshold to validate the presence of the watermark. The DWT based methods in [2, 3] also use the original image in a similar manner and are therefore not suitable for resolving rightful ownership.

So why use the original in the watermarking process at all? As pointed out earlier, robustness requires the watermark to be added in significant coefficients in the

*The support of the Office of Naval Research under grant N00014-96-1-0502 is greatfully acknowledged

0-8186-8821-1/98 $10.00 © 1998 IEEE

DCT domain. However, the order and number of these significant coefficients can change due to various image manipulations. Thus the original image is important in ascertaining the ordering of the watermark placed in the top coefficients of the image. In this paper we will show a way of getting around this "order sensitivity" issue without using the original during detection.

Among the methods which do not use the original for watermark detection, Piva *et al.* [4] suggested adding the watermark to a larger number of DCT coefficient which need not be significant. They order the DCT coefficients in a zig-zag scan and the *first* 16000 coefficients are left out. The watermark is added to the next 25000 coefficients. Watermark detection is performed by correlating these 25000 coefficients in the test image with the original copy of the watermark. Note that the original image is not required in this test. A larger number of coefficients is required here for a significant detector response as compared with the method in [1], since correlation is performed without subtracting out the original image. Since the watermark is added to such a large number of coefficients, visual masking is done in the spatial domain to prevent degradation in the perceptual quality of the image. But this masking cannot be taken into account in the process of watermark detection, leading to a comparatively poor detector response. Since the coefficients to which the watermark is added need not be significant, the watermark is susceptible to be removed by common image manipulations like Wiener filtering or compression with a low quality factor[8].

We present a method which can both add the watermark to the significant coefficients in the DWT domain *and* does not require the original image in the detection process. Since the watermark is added to significant coefficients in the DWT domain, our method is much more resistant to common image manipulations when compared with [4]. The amount of watermark added is adapted to the image so that less amount of watermark is added to a smooth image like *lena* and more to a not so smooth image like *baboon*. Further more, the time-frequency localization properties lead to implicit visual masking (as opposed to explicit visual masking in [4]); this improves the correlation coefficient in the detection process considerably. When compared with [4], our method results in significant amount of computational savings, since we cast a much smaller watermark and do not need to compute a visual mask.

2 The Proposed Method

Figure 1 shows a block diagram of the proposed method. We use a three level DWT with a Daubechies 8-tap filter. We leave out the low pass sub-band and pick all coefficients in the other sub-bands which are above a given threshold (T_1). Watermark is added to these coefficients only.

Although we add the watermark only to a few significant coefficients, an *image sized* watermark is being used. Thus the watermark at a particular location in the DWT of the image is fixed; there is no dependence on the order of the significant coefficients (which can change due to image manipulations) in the detection process. Since watermark detection involves finding the correlation coefficient, which is very sensitive to changes in the order of the vectors being correlated, order independence is a crucial factor in the success of the proposed method.

We add the watermark to all coefficients (barring the lowpass component) *above* a threshold T_1 rather than adding it to say the *top* 16000 coefficients. Smooth images like *lena* have much fewer number of coefficients above a threshold compared to an image like the *baboon*. Hence picking all coefficients above a threshold is a natural way of adapting the amount of watermark added to the image. Moreover it is found the small coefficients in the DWT domain are more susceptible to be corrupted by compression and other image manipulations like denoising as compared to the large coefficients.

Visual masking is implicit due to the time-frequency localization properties of the DWT. High pass bands, where the watermark is added, typically contain edge related information of the image. Furthermore, each coefficient in the high frequency bands affects only a spatially limited portion of the image. Thus, adding the watermark to significant coefficients in the high frequency bands is equivalent to adding the watermark to only the edge areas of the image, which makes the watermark invisible to the human visual system. This is corroborated in section 3

During watermark detection, we choose all the high pass coefficients above T_2 and correlate them with the original copy of the watermark. We use $T_2 = 50$ and $T_1 = 40$ (T_1 is the threshold used for watermark casting). $T_2 \geq T_1$ is necessary because we should not compute correlation over coefficients to which we have not added any watermark. We choose T_2 to be strictly larger than T_1 for robustness since some coefficients, which were originally below T_1, may become greater than T_1 due to image manipulations.

The equations used for watermark casting and detection are similar to those used in [4]:

$$V_i' = V_i + \alpha \mid V_i \mid x_i \tag{1}$$

where i runs over all DWT coefficients $> T_1$ (barring

the lowpass component). V_i denotes the corresponding DWT coefficient of the original image and V_i' denotes the DWT coefficient of the watermarked image. x_i is the watermark value at the position of V_i. x_i is generated from a uniform distribution of zero mean and unit variance. α is taken as 0.2.

For watermark detection the same procedure as above is followed but now only the coefficients (again barring lowpass component) $> T_2 > T_1$ are used as explained above. The correlation z between the DWT coefficients $\hat{V}$ of the corrupted watermarked image and a possibly different watermark Y is computed as

$$z = \frac{1}{M} \sum_i \hat{V}_i y_i \tag{2}$$

where i runs over all coefficients $> T_2 > T_1$ and M is the number of such coefficients.

The threshold S is defined as

$$S = \frac{\alpha}{2M} \sum_i | \hat{V}_i | \tag{3}$$

Piva *et. al.* [4] use a factor of 3M in the denominator of (3) instead of 2M. We have increased our threshold because the mutual correlation in our method is, in most cases, close to its theoretical value (because most of the high valued coefficients are preserved even after much image degradation and because we do not use any explicit visual masking) and therefore we can lend more fidelity to the method by increasing the threshold. Also, in our case, the number of samples over which the correlation is computed is typically smaller than in [4] and hence the cross-correlation may be higher.

3 Experimental Results

Figure 2 shows the original *lena* image, its watermarked copy and the watermark in the spatial domain. We see that the watermarked image is not distinguishable from the original image. Figure 2(c) shows that most of the watermark is added in edge regions of the image as claimed before. Thus there is no need for any explicit visual masking.

Figure 3(a) & (b) show the response with our method and with the method in [4] on the watermarked *lena* image. The dotted line shows the threshold S. We see that with our method the response is much stronger than its theoretically expected value ($2S$). For the method in [4], the response is close to its theoretically expected value ($3S$). With both methods, the cross-correlation is much lower than the mutual correlation.

We compressed the image with 5% quality JPEG compression (fig. 2(d)). Figures 3 (c) & (d) show that in our case the detector response is still well above the threshold, where as for [4], it falls below the threshold. In another experiment, the image was cropped (fig. 2(e)). Figures 3 (e) & (f) show that in our case the detector response is still well above the threshold; the scheme in [4] comes dangerously close to falling below the threshold.

Figures 3(g)-(i) show the effect of some other attacks on the image. We see that in all the cases, the mutual correlation is at least 1.5-2 times the threshold.

Acknowledgment

The first author would like to thank Prof. Kannan Ramachandran for letting him pursue digital image watermarking as term-project for the Wavelets in Signal Processing course.

References

[1] I. J. Cox, F. T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Transactions on Image Processing*, vol. 6, pp. 1673–1687, Dec. 1997.

[2] D. Kundur and D. Hatzinakos, "A robust digital image watermarking method using wavelet-based fusion," in *International Conference on Image Processing*, vol. III, pp. 544–547, 1997.

[3] X.-G. Xia, C. G. Boncelet, and G. R. Arce, "A multiresolution watermark for digital images," in *International Conference on Image Processing*, vol. III, pp. 548–551, 1997.

[4] A. Piva, M. Barni, F. Bartolini, and V. Cappellini, "DCT-based watermark recovering without restoring to the uncorrupted original image," in *International Conference on Image Processing*, vol. III, pp. 520–523, 1997.

[5] I. Pitas, "A method for signature casting of digital images," in *International Conference on Image Processing*, pp. 215–218, 1996.

[6] R. G. van Schyndel, A. Z. Tirkel, and C. F. Osborne, "A digital watermark," in *International Conference on Image Processing*, pp. 86–89, 1994.

[7] S. Craver, N. Memon, B.-L. Yeo, and M. M. Yeung, "Can invisible watermarks resolve rightful ownerships?," in *IBM Cyber Journal*, http://www.research.ibm.com:8080, July 1996.

[8] K. Ratakonda, R. Dugad, and N. Ahuja, "Digital image watermarking: Issues in resolving rightful ownership," in *International Conference on Image Processing*, (Chicago), Oct. 1998.

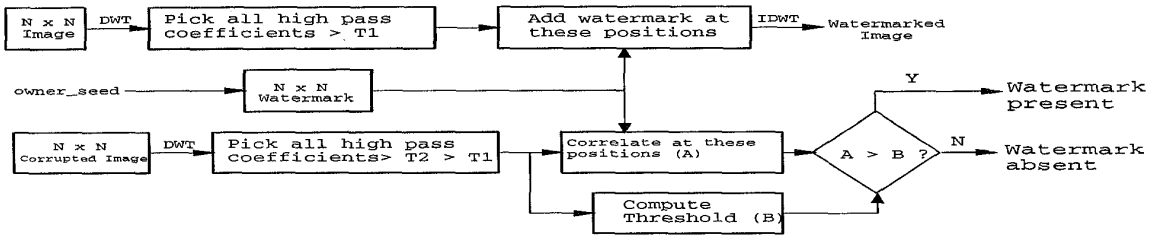

Figure 1: The proposed method: Top part shows watermark casting and bottom part shows watermark detection.

(a)original *lena* (b)*lena* watermarked (c)the watermark

(d)JPEG compressed (e)cropped (f)median filtered

Figure 2: (a) original *lena* image (b) *lena* watermarked with our method (c) The difference of (a) and (b) i.e. the watermark as it appears in the spatial domain. Note that most of the watermark gets added in edge regions where it is perceptually not visible. The watermarked image in (b) and the difference image in (c) have been scaled for display purpose. Watermarked image in (b) with (d) JPEG 5% quality compression, (e) cropping to retain only the central portion and (f) median filtering with a 5×5 window

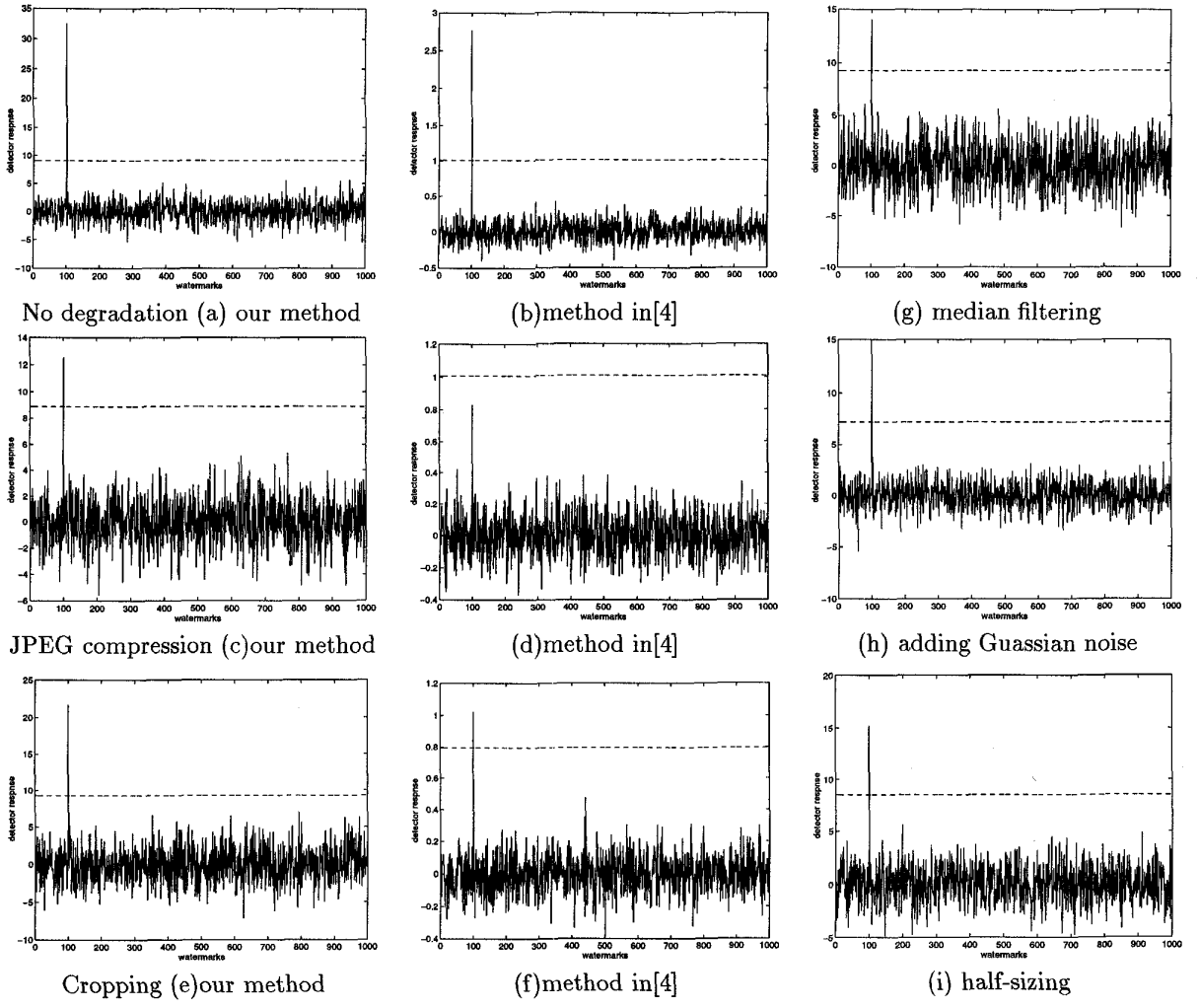

Figure 3: (a) and (b) show the responses on the watermarked images(fig. 2(b)) with our method and the method in [4] respectively. The graphs show the plot of different watermark seeds (x-axis) vs. their detector response (y-axis). The original image was watermarked with a seed of 100. (c) & (d) show the response after JPEG 5% quality compression (fig. 2(d)) and (e) & (f) after cropping (fig. 2(e)) the image watermarked with our method and the method in [4] respectively. Responses with our method after (g) median filtering with a window size of 5×5, (h) adding Gaussian noise with $\sigma^2 = 600$,(i) subsampling by two and resizing.