Marking and Detection of Text Documents Using Transform-domain Techniques

Yong Liu, Jonathon Mant, Edward Wong, Steven Low

Dept. of Electrical & Electronic Engineering
University of Melbourne
Parkville Vic 3052, Australia

Abstract

We investigate the use of frequency domain techniques to watermark text documents. A text image is essentially binary and hence contains large high-frequency components. This has several implications on the obtrusiveness and detection performance of frequency domain marking of text images, as illustrated by our extensive experiments. Generally marking is more obtrusive on a text than pictorial image. It almost always creates a ``dirty'' background. ``Cleaning'' the background by thresholding light greys to white renders the watermark less obtrusive but also sharply reduces the detector response, making it unrobust against noise. Both text and pictorial images seem very susceptible to shifting; this contrasts with extreme robustness against shifting of spatial domain marking through line or word shifting. Finally, we explore the combination of spatial domain marking and frequency domain detection and present preliminary experimental results on the combined approach.

Keywords: Watermarking, Transform-domain watermarking, Text documents, Detection

1 Introduction

This paper is motivated by the need to discourage illicit distribution of *formatted* text documents through watermarking. The mark must be indiscernible, yet it must survive common processes a document might be subjected to that introduce noise and distortions. A natural way to watermark a text document that exploits its regular structure, words and lines, is to embed the marks by shifting slightly some lines or words [2,5,6]; see also [9] for a proposed architecture for large scale distribution of text documents in a way that protects copyright and privacy. These techniques have the advantage of being unobtrusive and requiring minimal computation both for marking and for detection. Numerous watermarking methods have been proposed for pictorial and video images, see e.g., [8] and references therein. An important class of them mark an image by modifying the transform coefficients, e.g., [1,3,4,7,8]. They have been shown to be unobtrusive and robust against noise and distortions when applied to pictorial and video images. In this paper we explore its application to text image through extensive experiments.

In Section 2 we briefly review the original Cox et al. algorithm [3] with which we experiment as a representative transform domain technique. We present results that show the "dirty" background introduced by marking and the effect of cleaning on detection performance. Cleaning the background renders the watermark indiscernible but also sharply reduces the detector response, making it unrobust against noise and distortions. Motivated by the unobtrusiveness of spatial domain marking, we propose in Section 3 a combined algorithm that marks a text document by line or word shifting and detects the watermark using Cox et al. algorithm in the frequency domain. Preliminary experiments indicate that the approach is promising in offering a better detection performance for word shifting than the correlation detector of [5].

2 Transform Domain Marking of Text

In this section we apply the transform domain watermarking scheme of [3] to mark a text image. After a brief review of the scheme, we will present examples to illustrate the "dirty" background generally introduced by such a scheme. We will then consider the effect of "cleaning" the background on detection performance. We conclude that transform domain marking is usually more obtrusive on text than on pictorial images, and that cleaning the background after marking reduces both the obtrusiveness and, significantly, the detector response.

2.1 The DCT Algorithm

A watermark consists of a sequence of N real numbers, $X = (x_1, x_2, \ldots, x_N)$. The distribution of x_i is implementation specific but Cox et al. [3] recommend a continuous, normal distribution with zero mean and unit variance, which we use in our implementation. We have implemented the Cox et al. algorithm using MATLAB, v5.0 running on a Unix platform.
An image, pictorial or text, can be represented as:

$D(x,y) \varepsilon [0,1, \ldots, 255]$, where $x = 0, \ldots, W-1$; $y = 0, \ldots, H-1$

and its 2-dimensional, frequency domain transform as:

$C_D(k_x,k_y)=F[D(x,y)]$, where $k_x = 0, \ldots, W-1$; $k_y = 0, \ldots, H-1$

where W is the width of the image and H is the height. Following [3], we will use the 2-Dimensional Discrete Cosine Transform (DCT) in our implementation.

From $C_D(k_x,k_y)$, we extract a sequence of values, $V = (v_1, v_2, \ldots, v_N)$. The sequence V consists of the *largest* N elements of $C_D(k_x,k_y)$, excluding the DC component (located at $k_x=0$, $k_y=0$)). Into this sequence, the encoder embeds the watermark, X, to obtain a watermarked sequence of values, $V' = (v_1', v_2', \ldots, v_N')$. [3] describes various ways to compute v_i' from v_i and x_I, and we will use:

$$v_i' = v_i(1 + \alpha x_i)$$

V' is then inserted back into $C_D(k_x,k_y)$ in place of V to obtain a watermarked copy D'(x,y). This document will then undergo many processing operations to become $D^*(x,y)$. Given $D^*(x,y)$ and the location of the watermark, a detector can extract $V^*=(v_1^*, v_2^*, \ldots, v_n^*)$. Given D(x,y), it can then extract X^* using V^*-V, where X^* is a possibly corrupted watermark which is compared to X for statistical significance by measuring their similarity. Here large values are significant by the standard significance tests:

$$\mathrm{sim}(X,X^*) = \frac{X^* \cdot X}{\sqrt{X^* \cdot X^*}} \qquad (1)$$

It is important to ensure that X^* and X have the same mean lest the similarity metric be biased by large distortions in X^*. Thus, following [3], we zero the (sample) mean of each x_i^* prior to calculating sim(X,X*): $x_i^* \leftarrow x_i^* - \overline{X}^*$ where $\overline{X}^*$ is the sample mean.

The choice of N dictates the degree to which the watermark is spread out among the relevant components of the image.

2.2 Images under Test

Our tests are conducted on a text taken from Dickens's "A Tale of Two Cities" as shown in Figure 1. We also use the picture shown in Figure 2 as a control to compare between the marking and detection of text and pictorial images. The dimensions of each test image are standardized to those of the pictorial image, i.e., to 569x319 pixels.

It was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the epoch of belief, it was the epoch of incredulity, it was the season of Light, it was the season of Darkness, it was the spring of hope, it was the winter of despair, we had everything before us, we had nothing before us, we were all going direct to Heaven, we were all going direct the other way - in short, the period was so far like the present period, that some of its noisiest authorities insisted on its being received, for good or for evil, in the superlative

Figure 1: Original text image

Figure 2: Original pictorial image (control)

2.3 Obtrusiveness

Necessarily whether or not a watermark is obtrusive is subjective. Figures 3-8 show the result of marking on the pictorial and

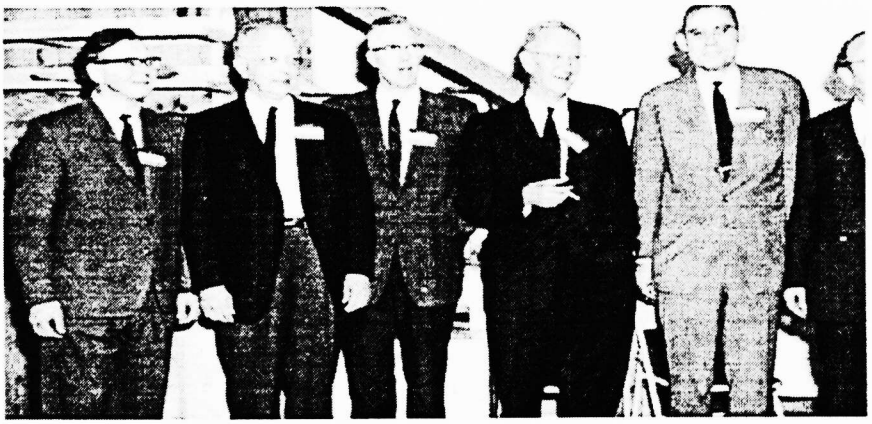

Figure 3: Pictorial image, $\alpha = 0.1$, $N = 10,000$

text images at different values of α and N. We have tested both images for a range of α and N values, and we now comment generally on the results.

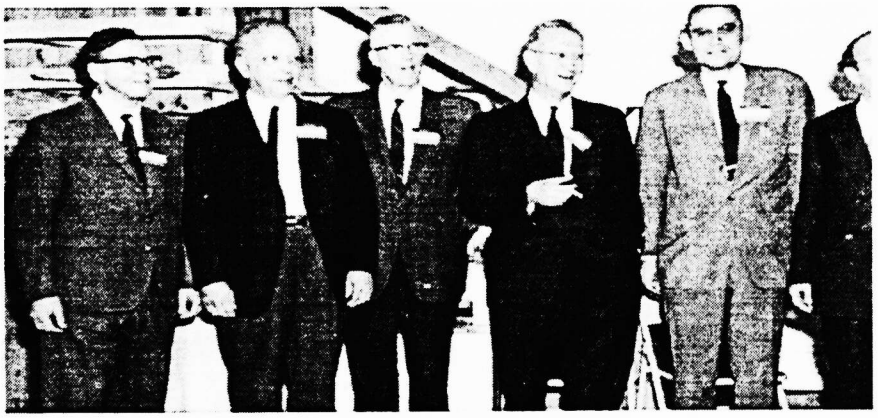

Figure 4: Pictorial image, $\alpha = 0.1$, $N = 50,000$

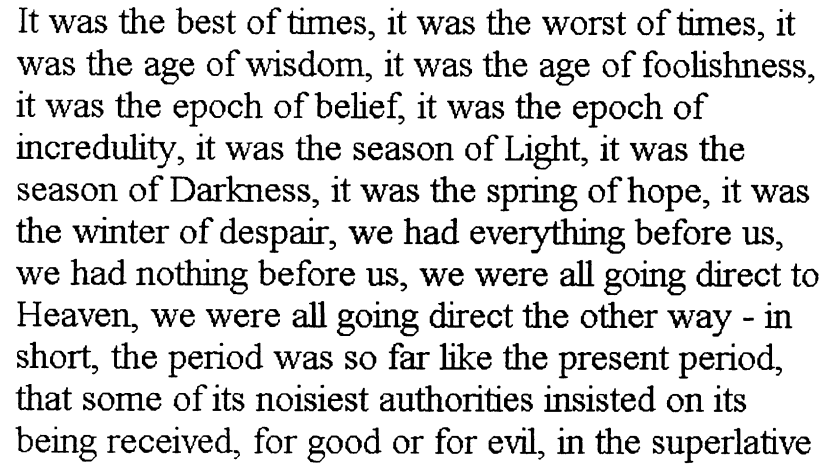

Figure 5: Text image, $\alpha = 0.1$, $N = 10{,}000$

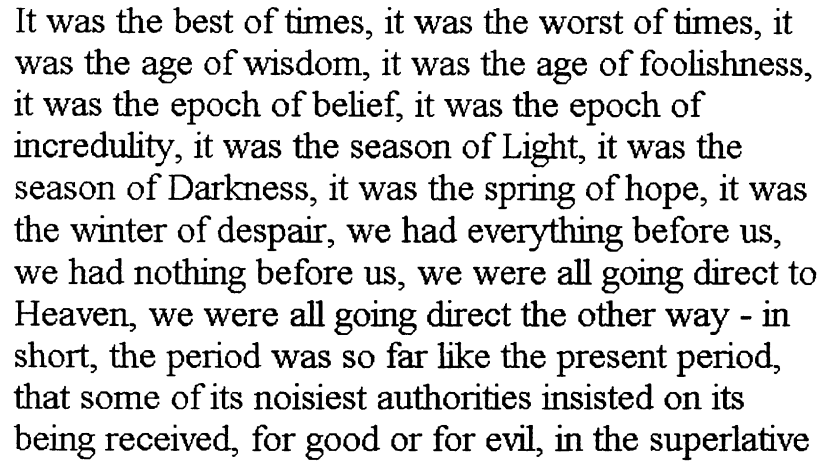

Figure 6: Text image, $\alpha = 0.1$, $N = 50{,}000$

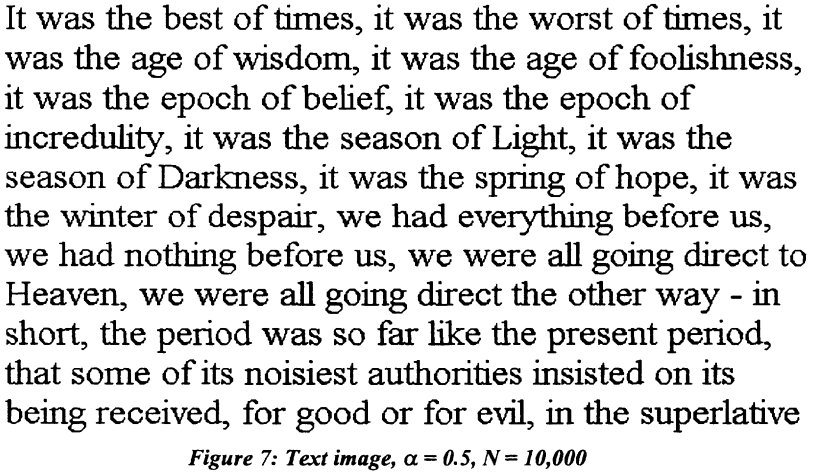

Figure 7: Text image, $\alpha = 0.5$, $N = 10{,}000$

It was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the epoch of belief, it was the epoch of incredulity, it was the season of Light, it was the season of Darkness, it was the spring of hope, it was the winter of despair, we had everything before us, we had nothing before us, we were all going direct to Heaven, we were all going direct the other way - in short, the period was so far like the present period, that some of its noisiest authorities insisted on its being received, for good or for evil, in the superlative

Figure 8: Text image, $\alpha = 1.0$, $N = 10{,}000$

Generally the watermark becomes more obtrusive as N or α increases. In particular, α much less than 0.5 should be used. This is natural since the larger the watermark N the more frequency bins altered, and the larger the value of α, the greater the change to the chosen frequency bins. Either leads to a greater modification in the spatial domain. Interestingly, marking of the text image seems more obtrusive, in two regards. First increasing N has a more substantial impact for the text image than pictorial image, and second the watermark becomes "obtrusive" for the text image at a much lower α value than for pictorial image. The first observation can be explained by the different characteristics of text and pictorial frequency spectra, and the second by human perception.

Due to its binary nature, a text image is rich in high frequencies. This contrasts with the frequency spectrum of a pictorial image which typically drops off rapidly at high frequencies. This is illustrated in Figures 9 which shows that energy concentrates only in low frequencies for the pictorial image but is roughly uniformly distributed across all frequencies for the text image. This implies that, for the pictorial image, once the initial set of most significant frequency components are marked, additional marks are insignificant. For the text image, however, adjustment made to each frequency bin has a consistent effect in the spatial domain.

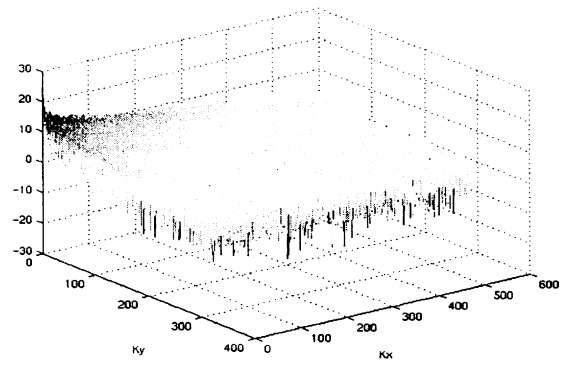

Figure 9 (a). 2D-DCT spectrum of pictorial image

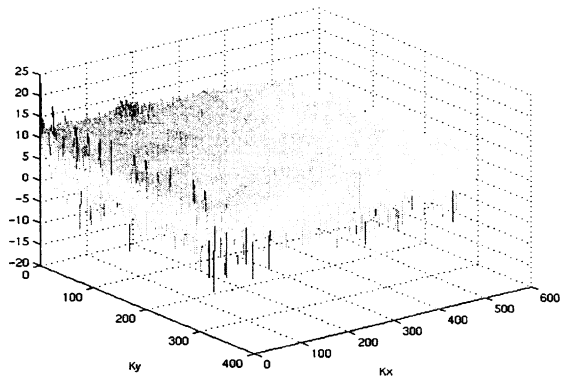

Figure 9 (b). 2D-DCT spectrum of text image

For text documents, the visually unappealing aspect of the watermark stems not so much from "light colored" speckling of the black foreground, but rather from the "dark colored" speckling of the white background. This is what we refer to as "dirty" background.

Figure 9 (c). Pictorial image, distribution of frequency bin marked

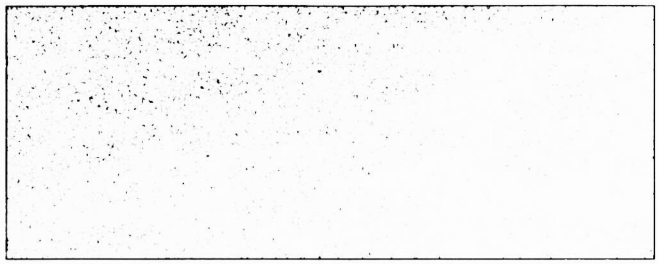

Figure 9 (d). 20-point text image, distribution of frequency bin marked

We have tried a few methods, including restricting marking to a particular area of the frequency spectrum, but, among them, the only method that substantially reduces the obtrusiveness is "cleaning" the background by thresholding a gray pixel to white. In the following subsections we examine the effect of cleaning on detector response.

2.4 Detector response with cleaning

Table 1 illustrates the detector resonse given by equation (1) as α is varied for the pictorial and text images with and without cleaning.

Range of α	Pictorial	Text (no cleaning)	Text (with cleaning)
0.1 – 0.5	100	96	55

Table 1: Detector response with various α

Hence we see that cleaning, while renders marking less discernible, also severely reduces the detector response, in this case by 43%. This diminishes the robustness of the watermark, as the following experiments show.

2.5 Robustness against noise

Many processes performed on documents can be simulated, in whole or part, by a noise attack. To test the robustness of the watermark to this attack a fraction of pixels are selected at random and their intensity is modified by an amount that is normally distributed according to $N(0,\sigma^2)$. We examine detector responses for varying percentages of the image marked (5%, 10%, 25%, 50% and 75%) and for varying σ (10, 20, ..., 100).

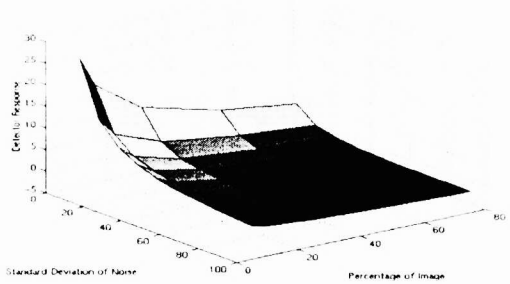

Figure 10 (a). Noise attack, α=0.1, pictorial image

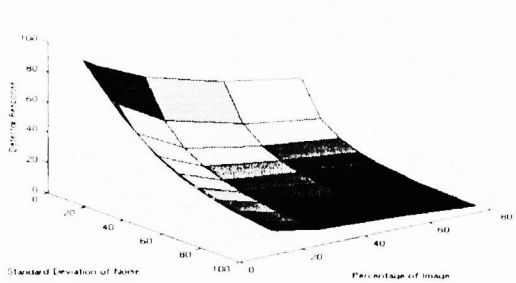

Figure 10 (b). Noise attack, α=0.1, text image without cleaning

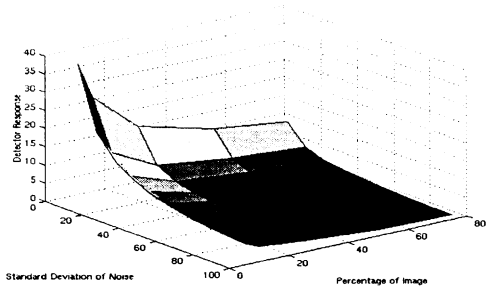

Figure 10 (c). Noise attack, α=0.1, text image with cleaning

Figure 10 demonstrates the effect of each of these parameters on detector response for the pictorial and text image with and without cleaning. From the figure we see that the watermark is significantly more robust in text than in a picture, but that noise rapidly reduces the detector response. Moreover cleaning reduces the robustness sharply, in many cases by 50% or more.

2.6 Robustness against cropping

Often only part of a document will be copied, although this still violates intellectual property rights. The watermark should therefore be discernible from only part of the document; that is, it must survive cropping. We implement detection by inserting the cropped portion of the document into the original *unmarked* copy to obtain $D^*(x,y)$ - the document upon which the detector operates. That is, if we have an *a* pixel x *b* pixel portion of a marked document based at the origin, then the corrupted copy is:

$$D^*(x,y) = \begin{cases} D'(x,y), & 0 < x < a,\ 0 < y < b \\ D(x,y), & \text{elsewhere} \end{cases}$$

We have used rectangular extracts of complete lines based at the origin with the cropped portion having a size from 569x50 to 569x300 pixel^2 (the entire image is 569x319 pixel^2). Figure 11 shows the detector response with varying α, with and without cleaning. As for pictorial images (see [3]), the watermark in a text image is very robust against cropping: it can be extracted from just a small portion of the image. Once again cleaning sharply reduces the robustness. Note also that the response is rather insensitive to α.

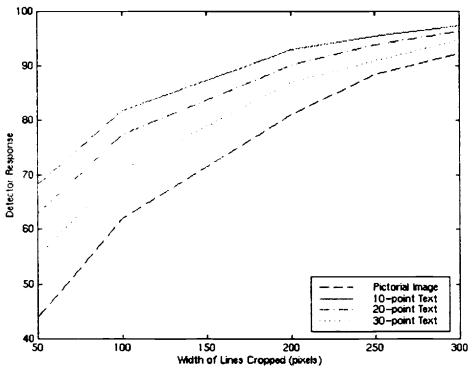

Figure 11(a). Robustness against cropping

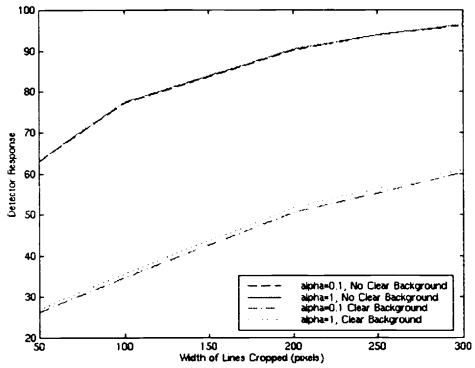

Figure 11(b). Robustness with and without cleaning

2.7 Robustness against translation

Scanning or photocopying a document often results in the document being slightly off centre: that is, in it being shifted by one or two pixels up, down, to the left or to the right. We test the robust of the watermark to such translation. That portion of the expected matrix size not filled by the image as a result of shifting is set to white. Table 2 charts the results.

	Down 1	Right 1	Up 1	Left 1	Down 2	Right 2	Up 2	Left 2
Pictorial image, α=0.1	3.2	1.9	3.5	2.8	1.2	-1.3	0.2	-0.3
20-point image, α=0,1, no cleaning	2.4	0.7	3.3	1.8	-1.0	-1.6	-0.3	-0.0
20-point image, α=1, no cleaning	13.4	12.4	11.8	13.9	2.5	-4.0	3.7	-2.7
20-point Image, α=0,1, with cleaning	1.1	-0.6	2.0	0.6	-1.3	-1.3	-0.7	0.2
20-point Image, α=1, with cleaning	2.4	0.9	0.2	2.5	-0.6	-1.6	-0.1	-0.2

Table 2: Detector responses after shifting

From the table we see that the watermark is not robust against translation for either pictorial or text image. Only with α=1, no cleaning, can it survive 1 pixel shifts, but as discussed earlier, the image is distorted beyond usability at such a large α value. Hence some form of compensation must be performed before detection is attempted. This contrasts with the extreme robustness against translation of the spatial domain marking and detection scheme in [2,6,5].

2.8 Binarizing

The previous results have revealed that by setting all pixels with an intensity below a certain threshold to white (i.e., cleaning), the strength of the watermark is attenuated. Binarizing goes one step further by setting all pixels with an intensity above the threshold to black. Since a text image is essentially binary, binarizing a watermarked image is expected to remove the watermark, as Figure 12 confirms. In contrast, the watermark in a pictorial image seems more robust against the attack. Moreover binarizing is much less obtrusive for text images than for pictorial images.

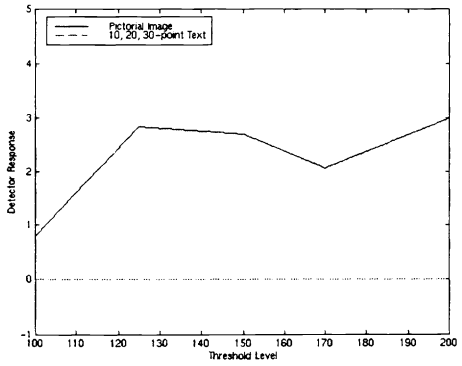

Figure 12. Binarizing destroys watermark in text image, α=0.1

3 A Combined Approach

Several spatial domain marking techniques have been proposed in [2,5,6] for formatted text, where the watermark is embedded by shifting lines or words by 1 or 2 pixels. These techniques have the advantage of being unobtrusive. Moreover, at least for line shifting, it is extremely robust against severe noise and distortions introduced in common document processing such as printing, photocopying and facsimile transmission; see the experimental results in [2,6] and a performance comparison between line and word shifting in [5]. Table 3 summarizes the comparison between transform domain and spatial domain algorithms for text marking.

	Cox, et al. Algorithm	Line Shifting	Word Shifting
Obtrusiveness	Bad	Excellent	Excellent
Noise resistance	Excellent	Excellent	Bad
Cropping	Excellent	Bad	Bad

Translation	Bad	Excellent	Good
Binarizing	Bad	Excellent	Excellent
Bit rate	Excellent	Bad	Good
Print, xerox, scan	Good	Excellent	Bad
Computation	Bad	Excellent	Excellent

Table 3: Comparison of transform and spatial domain marking

Compared with word shifting, transform domain marking is much more robust against noise and cropping, but is obtrusive. This motivates the approach that marks a text document by line or word shifting and detects the watermark using Cox et al. algorithm. It attempts to combine the unobtrusiveness of spatial domain techniques and the good detection performance of frequency domain techniques.

3.1 Algorithm

Marking
Apply line or word shifting and store the spatial watermark X_s in a user profile database.

Detection
1. Apply the spatial watermark X_s to the original document D(x, y) to obtain the marked document D'(x, y).
2. Compute the transform of the original and marked document:
   $C_D(k_x, k_y) = F[D(x, y)]$ and $C_{D'}(k_x, k_y) = F[D'(x, y)]$.
3. Extract the frequency watermark $X = (x_1, \ldots, x_N)$ where X consists of the *largest* N values of the absolute differences in $|C_{D'}(k_x, k_y) - C_D(k_x, k_y)|$.
4. Compute the transform of the corrupted document $C_{D^*}(k_x, k_y) = F[D^*(x, y)]$ and extract the corrupted frequency watermark $X^* = (x^*_1, \ldots, x^*_N)$. Here X^* consists of the N values of the absolute differences in $|C_{D^*}(k_x, k_y) - C_D(k_x, k_y)|$ corresponding to the frequency bins selected in Step 3.
5. Zero the (sample) means of X and X*: $x_i \leftarrow x_i - \overline{X}$ and $x^*_i \leftarrow x^*_i - \overline{X}^*$.
6. Compute sim(X, X^*).

It is interesting to contrast the modification to the frequency spectrum using Cox et al. algorithm and line or word shifting. Figure 13 shows the spectral distribution of the 10,000 watermark elements for (a) Cox et al. algorithm, for (b) line shifting, and for (c) word shifting.

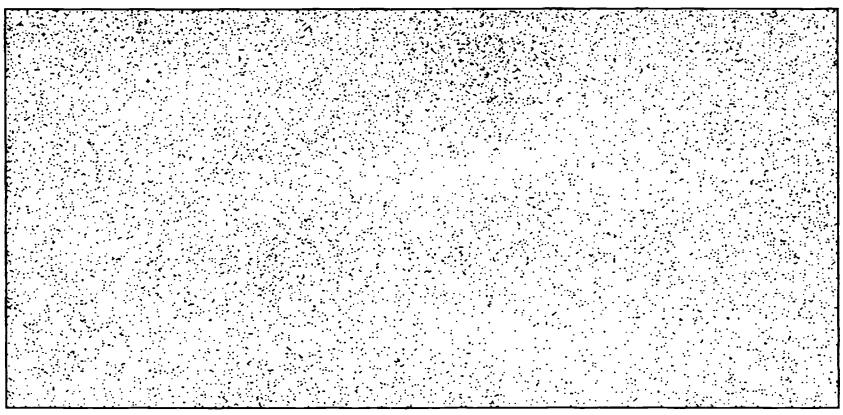

Figure 13(a). Spectral distribution of watermark, Cox et al. algorithm

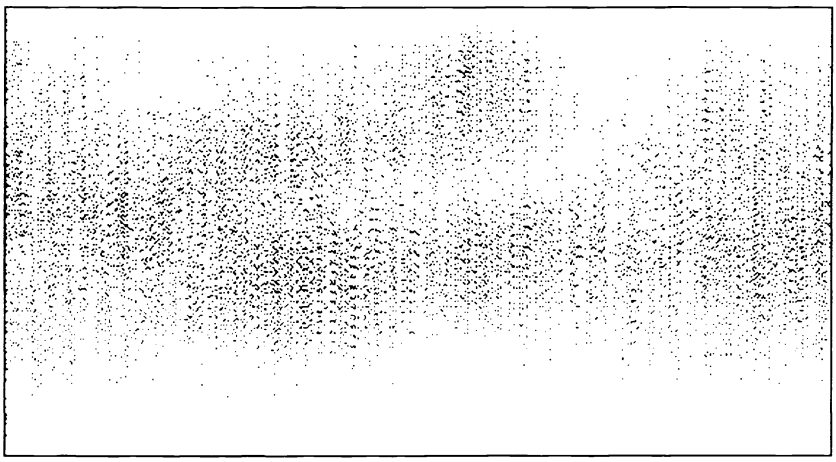

Figure 13(b). Spectral distribution of watermark, line shifting

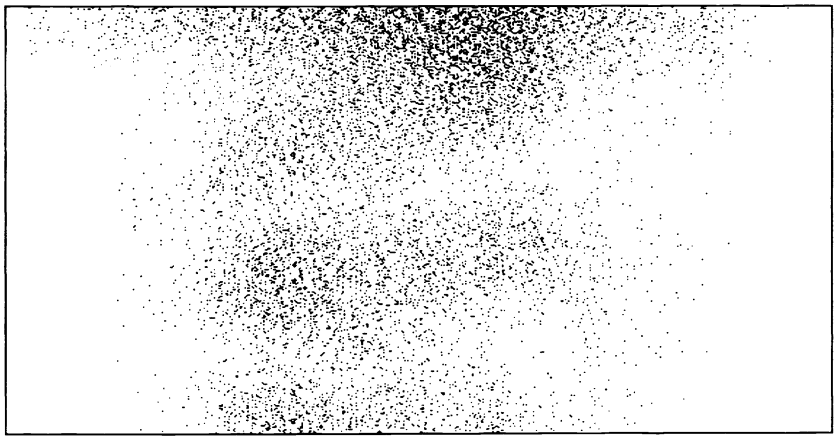

Figure 13(c). Spectral distribution of watermark, word shifting

3.2 Robustness against false alarm

As line or word shifting can accommodate much fewer marks than frequency domain watermarking it is critical to determine the minimum hamming distance between valid watermarks so that false alarm can be kept sufficiently unlikely. Figure 14 illustrates the detector response as a function of the hamming distance between the actual and false spatial watermarks. We see that, for word shifting, a hamming distance of 10 is required for the detector response with false watermark to be no more than 50% of that with the actual watermark, and for line shifting, a hamming distance of 3 is required.

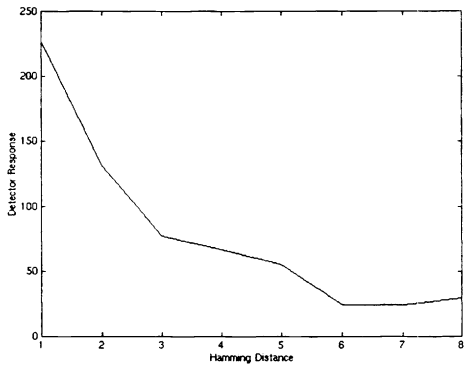

Figure 14(a). Detector response, line shifting

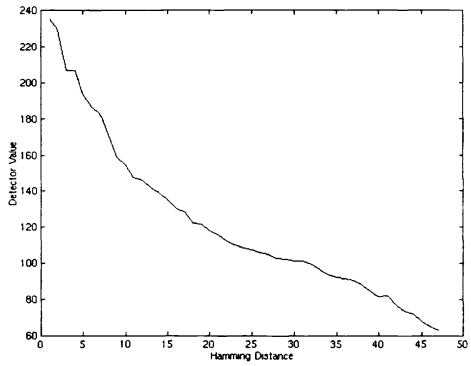

Figure 14(b). Detector response, word shifting

3.3 Robustness against cropping

We have repeated the test of Section 2.6 on cropping. The detector response as a function of the size of the cropped portion, for both line and word shifting, is shown in the following table. Again the entire image is 569x319 pixel^2.

	569x50 extract	569x100 extract	569x200 extract	569x250 extract	569x300 extract
Line shifting	34	51	100	136	207
Word shifting	59	77	110	131	163

Table 4: Robustness against cropping

4 Conclusion

In this paper we have studied frequency domain watermarking of text images using Cox et al. algorithm. The dirty background often introduced by such techniques renders the watermark obtrusive. We have presented experimental results that illustrate that, while cleaning the background is effective in making the watermark indiscernible, it also sharply reduces the detection robustness against noise and distortions. Motivated by the unobtrusiveness of spatial domain watermarking, we have proposed a combined approach that marks a text document by line or word shifting and detects the watermark in the frequency domain using Cox et al. algorithm. Preliminary experimental results illustrate the effectiveness of this approach.

References

[1] W. Bender, D. Gruhl and N. Morimoto, "Techniques for data hiding," in *Proc. SPIE*, Feb. 1991, pp. 2420-2440.

[2] J. Brassil, S. Low, N. Maxemchuk and L. O'Gorman, "Electronic marking and identification techniques to discourage document copying," *IEEE J. Select. Areas Commun.*, Vol. 13, pp. 1495-1504, Oct. 1995.

[3] I. Cox, J. Kilian, T. Leighton and T. Shamoon, "Secure spread spectrum watermarking for multimedia," in *Proc. First Int. Workshop Information Hiding*, R. Anderson, Ed., Cambridge, U. K.: Springer-Verlag, May/June 1996, pp. 183-206.

[4] E. Koch and J. Zhau, "Toward robust and hidden image copyright labeling," in *Proc. 1995 IEEE Nonlinear Signal Processing Workshop*, 1995, pp. 452-455.

[5] S. Low and N. Maxemchuk, "Performance comparison of two text marking methods," *IEEE J. Select. Areas Commun.* Vol. 16, pp. 561-572, May 1998.

[6] S. Low, N. Maxemchuk and A. Lepone, "Document identification for copyright protection using centroid detection," IEEE Trans. on Commun., Vol. 46, pp. 372-383, March 1998.
[7] M. D. Swanson, B. Zhu and A. H. Tewfik, "Multiresolution scene-based video watermarking using perceptual models," *IEEE J. Select. Areas Commun.*, Vol. 16, pp. 540-550, May 1998.
[8] Special Issue on Copyright and Privacy Protection, *IEEE J. Select. Areas Commun.*, Vol. 16, May 1998.
[9] J. Brassil, S. Low and N. Maxemchuk, "Copyright protection for the electronic distribution of text documents", IEEE Proceedings, to appear 1999.