Performance Comparison of Two Text Marking Methods

Steven H. Low, *Member, IEEE*, and Nicholas F. Maxemchuk, *Fellow, IEEE*

***Abstract*—A text document typically consists of a collection of regular structures such as words, lines, and paragraphs, a slight movement of which seems less perceptible than, say, dithering of the document image. In this paper we exploit this property to watermark formatted text documents by shifting slightly certain lines and words in order to discourage illicit distribution. We analyze two methods for reliable document identification in the presence of severe distortions introduced by photocopying, facsimile transmission, and other processing. The correlation method uses document profiles directly for detection. To eliminate the effect of certain distortions, the centroid method bases its decision on the distances between the centroids of adjacent profile blocks. We present the maximum likelihood detectors for both methods and evaluate their relative performance. Our analysis indicates that line-shift generally has a smaller error than word-shift detection, and that the correlation detector outperforms the centroid detector provided certain distortions can be accurately compensated for before detection is attempted. These results have been applied to implement a marking and identification system and preliminary experimental results have been very promising.**

***Index Terms*—Centroid detection, correlation detection, detection performance, document marking.**

I. Introduction

A WAY TO discourage illicit reproduction of copyrighted or sensitive documents is to watermark the document before distribution. We have prototyped such a document marking and identification system. It automatically puts a unique and indiscernible mark on each document copy and registers its recipient. If an illicit copy is recovered, the system detects the mark from the copy, identifying the original recipient. This paper explains the detection methods used in the identification subsystem and analyzes their performance. Preliminary experimental results show that very reliable identification can be achieved in the presence of severe distortions introduced by photocopying, facsimile transmission, and other digital processing; see [11].

A text document typically consists of a collection of regular structures such as words, lines, and paragraphs. The basic idea of our approach is to exploit these regular structures in hiding marks. Since a slight movement of an entire structure seems less perceptible to human eye than, say, dithering of the document image, we can hide enough "signal strength" in such movements to achieve accurate detection without the marks being perceptible. For instance, a text line can be moved up to encode a "1" or down to encode a "0" by 1 pixel, or 1/300th inch, at 300 dot-per-inch (dpi) resolution.

This contrasts with other interesting image marking approaches [22], [1], [15], [24], [9], [10], [5], [14], [7], [21], [20], [8] that do not exploit the regular structure of text documents. The two approaches may be combined to encode more information: general image marking techniques can be applied to a document image that has already been marked by our techniques (but see comment below on binarization attack on marking of text images).

To mark a page, certain text lines are shifted slightly up or down from their normal positions or certain words are shifted slightly left or right. The shifting pattern is different on different copies. To detect the marking, the horizontal profiles of lines and vertical profiles of words are compiled from a digitized image of the page. We have experimented extensively with two detection methods. The first method, the correlation detection, treats a profile as a discrete time signal and chooses the direction of shift that is most likely to account for the observed corrupted signal. To eliminate the effect of certain distortions, the second method, the centroid detection, does not base its decision on the profile directly. It detects the marking from changes in the spacing between the centroids of profile blocks. For both methods, we have derived the maximum likelihood decision rules that minimize the average probability of error when all marking patterns are equally likely *a priori*.

The centroid method is explained in detail in [13]. In this paper we present the correlation method and compare the relative performance of the two methods in detecting line shifts and word shifts. The performance comparison, confirmed by experimental results, leads to our adopting the centroid method to detect line shifts and the correlation method to detect word shifts in our document marking and identification system.

In Section II, we define formally a profile and propose a simple noise model to model how a horizontal or vertical profile is corrupted by printing, photocopying, scanning, and other processing. Based on this profile model, we present in Section III the correlation and the centroid detectors and their probability of error. In Section IV, we use the error probabilities to make four performance comparisons. For each detection method we compare the performance of line versus word-shift detection, and for each type of profile we compare the performance of the centroid versus correlation detector. We found that line-shift detection generally enjoys a smaller error probability than word-shift detection, and that the correlation outperforms the centroid detector on typical profiles

Manuscript received March 1997; revised July 1997.

S. H. Low is with Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, VIC 3052, Australia (e-mail: slow@ee.mu.oz.au).

N. F. Maxemchuk is with AT&T Laboratories, Research, Florham Park, NJ 07932 USA (e-mail: nfm@research.att.com).

Publisher Item Identifier S 0733-8716(98)01108-1.

provided certain distortions can be accurately compensated for before detection is attempted. Finally, we describe briefly in Section V how these results are applied to implement a marking and identification system.

Watermarking as a means to protect copyright has received much attention recently. Several methods have been proposed to discourage illicit reproduction of picture and video images in [22], [1], [15], [24], [9], [10], [5], [14], [7], [21] and [20]. These general techniques either are not directly applicable to or do not exploit the regular structure of text documents. Moreover, while transform-based techniques seem ideal for marking images with rich greyscale, it may not be well suited for binary images, such as an text image, because slight perturbation of image intensity can be easily removed by binarization. In [6], a cryptographic system for the secure distribution of electronic documents is described. In [3], the approach to indiscernibly mark each document copy by varying the line or word spacing or by varying certain character features slightly is proposed. In [12], an experiment is reported that reveals that a document can be distorted much more severely in one direction than the other, and a marking and identification strategy that exploits this difference is described. The detection schemes reported in this paper are more sophisticated than those in [3] and [12]. In [2], several ways to assign unique identifiers to copies of digital data are studied that are secure against collusion among recipients to detect and remove the marking.

Finally we comment on the applicability of the proposed technique, suitable only for *formatted* text documents. Marks placed in a text, using *any* technique, including the proposed one, can always be removed by retyping the document. A large part of this effort may be automated by character recognition devices. Alternatively the marks can be concealed by dithering the positions that contain information by larger amounts than the encoder uses to enter the information. In contrast, marks placed in pictures or speech are assumed to be indelible. The ability to remove text marks limits its applications. Text marking is well suited for protecting modestly priced documents such as newspaper or magazine articles. We assume that if legal and illegal copies are distinguishable (a document with markings altered or removed can be easily identified to be illicit), and legal copies are affordable, then most people will not seek out illegal copies. A similar assumption is made by the cable TV industry, where viewers can either buy a device to unscramble the signal on a premium channel or pay the cable operator. Attacks on the proposed text marking method are further elaborated in [4]. Countermeasures can be devised to make the distortion needed to conceal marks intolerable, to make it difficult to forge valid marks, and to make the marks more difficult to remove. For example, a publisher may watermark a document in postscript, but distribute marked copies in bitmap or paper. Then the marking process takes much less time than applying typical image marking techniques on bitmap images of the text, and can be performed in real-time before distribution. Moreover, for the recipients, it will be difficult to remove the marks and more expensive to redistribute the illicit copies.

Throughout the paper $h(y)$ denotes an original unmarked and uncorrupted profile and $g(y)$ denotes its corrupted copy, marked or unmarked. By "$X := Y$" or "$Y =: X$," we mean "X is defined as Y."

II. Model

A. Profiles and Marking

Upon digitization the image of a page is represented by a function

$$f(x,y) \in [0,1], \qquad x = 0,1,\cdots,W, \qquad y = 0,1,\cdots,L$$

that represents the grayscale at position (x,y). Here, W and L, whose values depend on the scanning resolution, are the width and length of the page, respectively. The image of a text line is simply the function restricted to the region of the text line

$$f(x,y) \in [0,1], \qquad x = 0,1,\cdots,W, \qquad y = t,t+1,\cdots,b$$

where t and b are the top and bottom "boundaries" of the text line, respectively. For instance, we may take t or b to be the mid-point of the interline spacing. The *horizontal profile* of the text line

$$h(y) = \sum_{x=0}^{W} f(x,y), \qquad y = t,t+1,\cdots,b$$

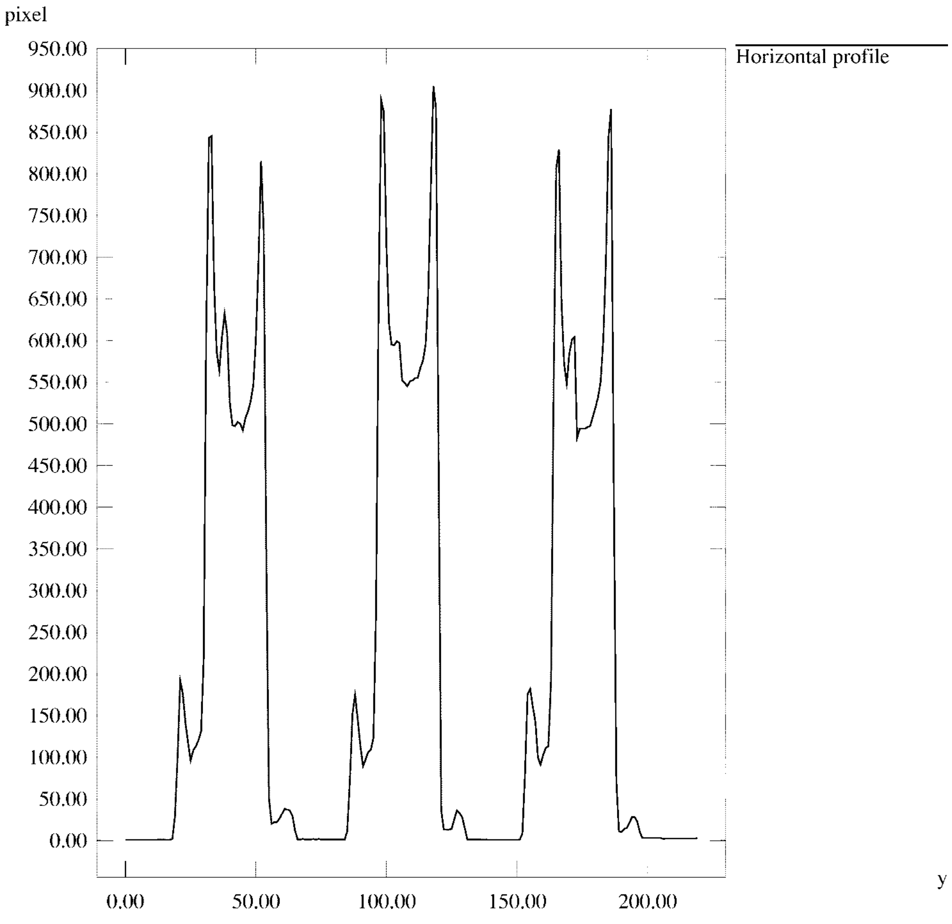
is the sum of grayscale along the horizontal scan-lines y. The *vertical profile* of the text line

$$v(x) = \sum_{y=t}^{b} f(x,y), \qquad x = 0,1,\cdots,W$$

is the sum of grayscale along the vertical scan-lines x. For simplicity we assume that $f(x,y)$, and hence the profiles $h(y)$ and $v(x)$, take continuous values.

Fig. 1 shows a typical horizontal profile of three text lines and a typical vertical profile of six words. Note the different scales on the two profiles. A horizontal profile consists of distinct "columns" and "valleys." The "columns" correspond to text lines and the "valleys" to interline spaces. The bulk of a column is several hundred bits for the shown digitization resolution. On the other hand, a vertical profile has shorter columns and narrower valleys that are much less distinguishable. These examples will be used for illustration throughout this paper.

A text line can be marked vertically by shifting it slightly up or down from its normal position to carry one bit of the copy's unique identifier. To compensate for major distortions, a line is marked only if it and its two neighboring lines are all sufficiently long. The neighboring lines, called the control lines, are not marked. Alternatively, a line can be marked horizontally by shifting certain words slightly left or right from their normal positions. The line is divided into some odd number of groups of words such that each group contains a sufficient number of characters. Each even group is then shifted, possibly *independently* of other even groups, while each odd group, called the control group, remains stationary. Hence multiple bits of information can be embedded in a line by word shifting. The control lines and control groups are used



(a)

Fig. 1. (a) Horizontal profile (resolution = 300 dots-per-inch).

to estimate and compensate for distortions in the horizontal profile and the vertical profile, respectively.

Both line-shift and word-shift marking can be considered within the same model where we have a profile, denoted by $h(y)$, that covers three "blocks"; see Fig. 5. For line-shift detection, each block is the horizontal profile of a text line. For word-shift detection, each block is the vertical profile of a group of words. The middle block is shifted slightly while the other two blocks, called the control blocks, are stationary.

B. Profile Noise

When the marked original is printed, photocopied, and then scanned, the text is typically distorted by translation, scaling, speckles (salt-and-pepper noise), rotation (skewing), blurring, and other random distortions. For example, a skew angle between −3° and +3° and an expansion or shrinkage of up to 2% have been observed in our experiments. From experience, photocopying introduces the most noise. A sample of an original text and its tenth copy is shown in Fig. 2.

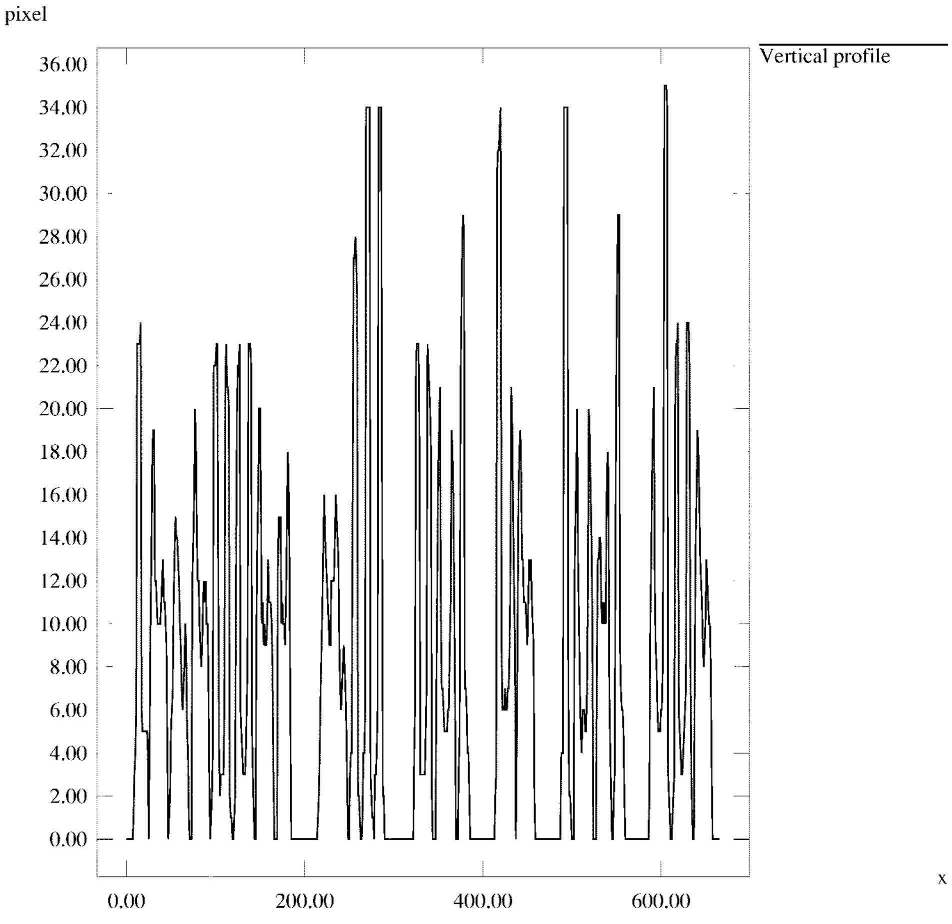
Before document profiles are compiled, the scanned image is first processed by standard document image processing techniques [18, Ch. 4], [16], [17] to remove skewing and speckles. Then profiles are compiled from the processed image. We assume that the translation and scaling are unknown but vary slowly with respect to the distance of encoding of a bit so that they are uniform across the encoding of a bit. They are estimated using the left and right control blocks and compensated for before detection is attempted; some heuristic schemes that have been tried are given in [12].

This series of processing is the motivation for us to include control blocks. The major distortions effect the marked blocks and the control blocks in a similar fashion. This is exploited to remove structural distortions on the marked blocks after estimating them from the control blocks. Furthermore, by estimating the correlation structure of the remaining noise on the control blocks, the remaining noise on the marked blocks can be whitened to a significant extent.

Hence we assume that a profile $h(y)$ on some interval $[b, e]$ after distortion compensation is corrupted only by additive noise $N(y)$ to become

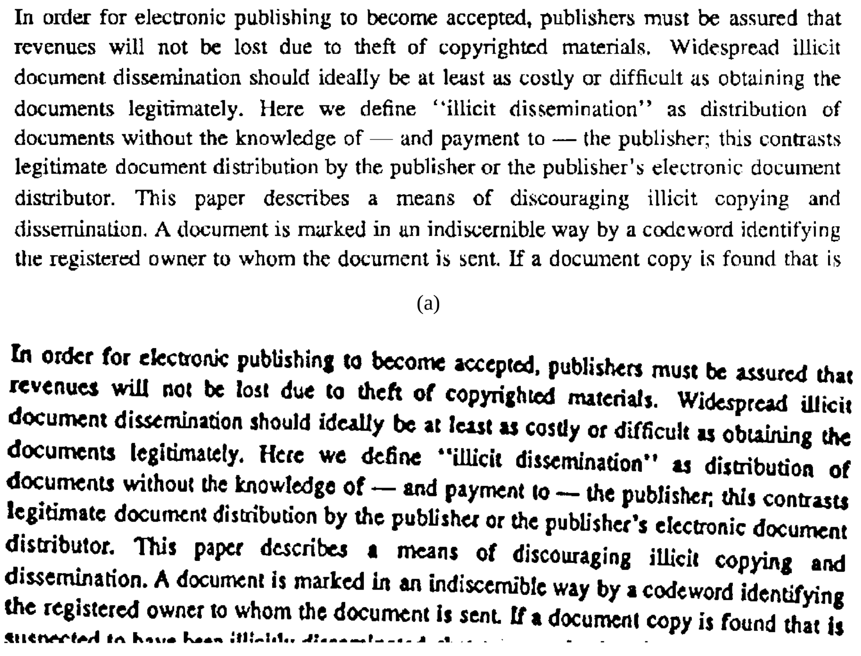
$$g(y) = h(y) + N(y), \qquad y = b, \cdots, e. \tag{1}$$

We assume that $N(y)$ are independent and identically distributed (i.i.d.) Gaussian random variables with mean 0 and variance σ^2. This white Gaussian noise models all the distortions not accounted for, as well as errors introduced in the compensation. A sample of noise $N(y)$ measured from a horizontal and a vertical profiles is shown in Fig. 3. The corresponding empirical distributions of $N(y)$ is shown in Fig. 4. From these figures, the Gaussian model seems reasonable as a first approximation.



(b)

Fig. 1. *(Continued.)* (b) Vertical (lower) profile (resolution = 300 dots-per-inch).



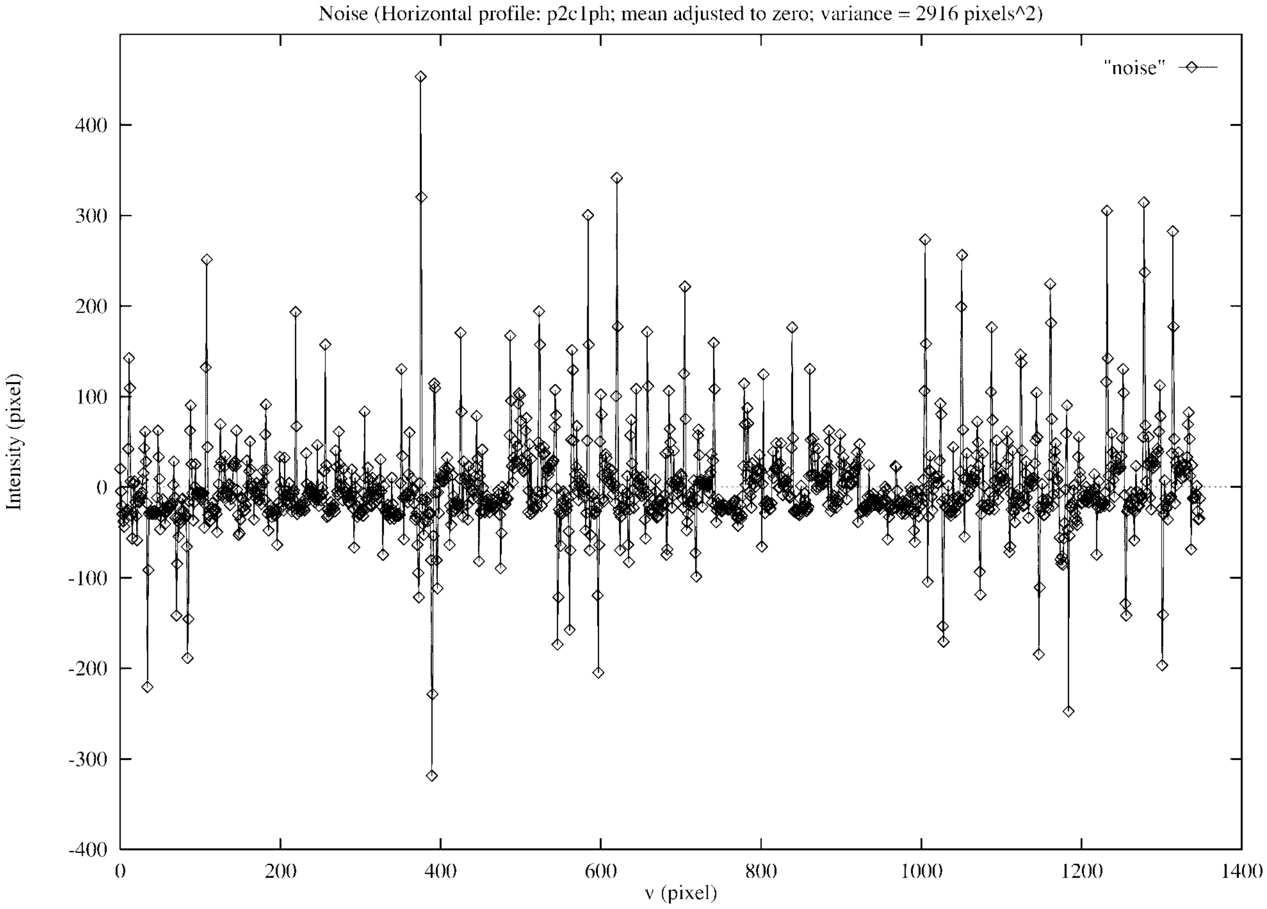
In order for electronic publishing to become accepted, publishers must be assured that revenues will not be lost due to theft of copyrighted materials. Widespread illicit document dissemination should ideally be at least as costly or difficult as obtaining the documents legitimately. Here we define "illicit dissemination" as distribution of documents without the knowledge of — and payment to — the publisher; this contrasts legitimate document distribution by the publisher or the publisher's electronic document distributor. This paper describes a means of discouraging illicit copying and dissemination. A document is marked in an indiscernible way by a codeword identifying the registered owner to whom the document is sent. If a document copy is found that is

(a)

In order for electronic publishing to become accepted, publishers must be assured that revenues will not be lost due to theft of copyrighted materials. Widespread illicit document dissemination should ideally be at least as costly or difficult as obtaining the documents legitimately. Here we define "illicit dissemination" as distribution of documents without the knowledge of — and payment to — the publisher; this contrasts legitimate document distribution by the publisher or the publisher's electronic document distributor. This paper describes a means of discouraging illicit copying and dissemination. A document is marked in an indiscernible way by a codeword identifying the registered owner to whom the document is sent. If a document copy is found that is suspected to have been illicitly disseminated

(b)

Fig. 2. Sample of an original text image (upper) and its tenth copy (lower).



(a)

Fig. 3. Sample profile noise measured from a (a) horizontal profile.

III. Detection and Performance

In this section we present the maximum likelihood decision rule and the probability of error for the correlation and the centroid methods.

A. Correlation Detection

Suppose we are given an original unmarked profile $h(y)$ and its noisy marked copy $g(y)$, each consisting of three blocks. In this subsection we present the correlation detector.

We think of the original unmarked profile $h(y)$ as a noisy communication channel, and marking as a signal that is transmitted onto this channel. Our objective is to detect the transmitted signal from the noisy received copy $g(y)$.

Suppose the original unmarked profile $h(y)$ has three blocks defined by the intervals $[b_1, e_1], [b_2, e_2]$, and $[b_3, e_3]$ as shown in Fig. 5. We assume that $h(y) = 0$ between these intervals. Let $h^l(y)$ be the resultant profile when the middle block is left shifted by $\epsilon > 0$

$$h^l(y) = \begin{cases} h(y), & y < b_2 - \epsilon \text{ or } y > e_2 \\ h(y+\epsilon), & b_2 - \epsilon \le y \le e_2 - \epsilon \\ 0, & e_2 - \epsilon < y \le e_2 \end{cases} \tag{2}$$

and $h^r(y)$ be that when the middle block is right shifted

$$h^r(y) = \begin{cases} h(y), & y < b_2 \text{ or } y > e_2 + \epsilon \\ 0, & b_2 \le y < b_2 + \epsilon \\ h(y-\epsilon), & b_2 + \epsilon \le y \le e_2 + \epsilon. \end{cases} \tag{3}$$

Naturally we assume the shift ϵ is smaller than the interblock spacing. The profile $g(y)$ compiled from the illicit copy and after distortion compensation is corrupted by additive white Gaussian noise such that

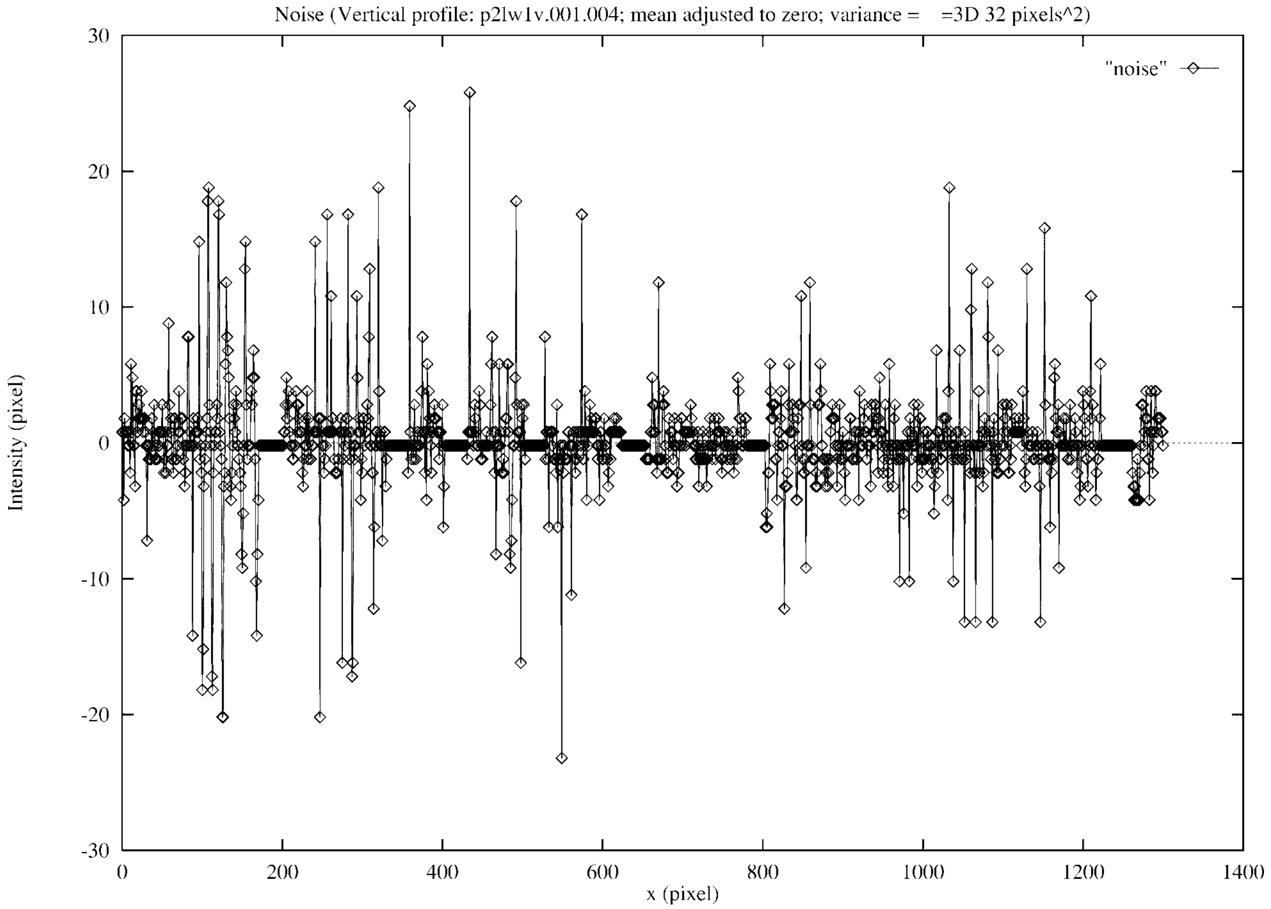
$$g(y) = h^l(y) + N(y), \qquad y = b_1, \cdots, e_3 \tag{4}$$

if the middle block is left shifted, and

$$g(y) = h^r(y) + N(y), \qquad y = b_1, \cdots, e_3 \tag{5}$$

if it is right shifted.

We have to decide whether the middle block is left or right shifted based on the observed profile $g(y)$. It is well known that the maximum likelihood decision rule under additive Gaussian noise can be implemented by a correlation detector [23, Ch. 4]. Standard procedure leads to the following propositions whose proofs are omitted.



(b)

Fig. 3. *(Continued.)* Sample profile noise measured from a (b) vertical profile.

Proposition 1: The maximum-likelihood decision given the observed profile $g(y)$ is

$$\begin{array}{ll} \text{left shift} & \text{if } \sum_{b_2}^{e_2} h(y)(g(y-\epsilon)-g(y+\epsilon)) \geq 0 \\ \text{right shift} & \text{otherwise.} \end{array}$$

Note that detection uses the profiles only around the middle block $[b_2, e_2]$.

We use the average probability of error

$$P_E := \tfrac{1}{2}\,(P(\text{decide left shift}|\epsilon) + P(\text{decide right shift}|-\epsilon)) \qquad (6)$$

to evaluate the performance of the decision rule, where $P(\text{decide left shift}|\epsilon)$ and $P(\text{decide right shift}|-\epsilon)$ are the probabilities of a wrong decision when the middle block is shifted right and left, respectively.

Proposition 2: The error probability of the maximum likelihood detector in Proposition 1 is

$$P_E = \operatorname{erf}\left(-\sqrt{\frac{\sum h^2(y) - \sum h^l(y)h^r(y)}{2\sigma^2}}\right)$$

where $\operatorname{erf}(x) := (2\pi)^{-1/2}\int_{-\infty}^{x} e^{-y^2/2}\,dy$.

B. Centroid Detection

We again are given an original unmarked profile $h(y)$ and its noisy marked copy $g(y)$, each consisting of three blocks. Each block represents a line in a horizontal profile or a group of words in a vertical profile. The centroid detection uses the distances between centroids of adjacent blocks as a basis for decision. It works well only if each block of the given profile can be accurately delineated. Assume this has been done.

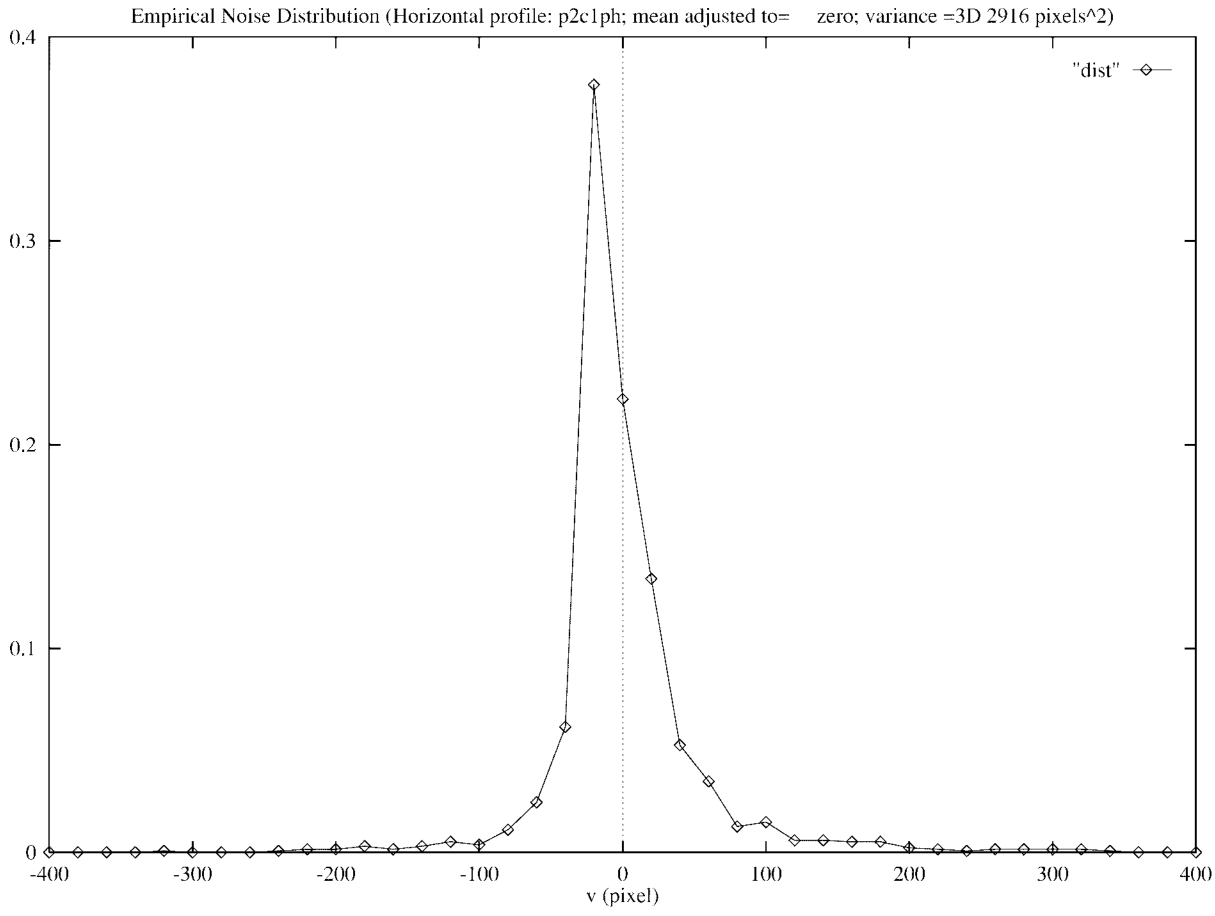
The original unmarked profile $h(y)$ consists of three blocks defined by the intervals $[b_1, e_1]$, $[b_2, e_2]$, and $[b_3, e_3]$, as shown in Fig. 5. The centroid of block $i, i = 1, 2, 3,$ is defined as

$$c_i = \frac{\sum_{b_i}^{e_i} y h(y)}{\sum_{b_i}^{e_i} h(y)}.$$

The profile compiled from the illicit copy and after distortion compensation is

$$g(y) = h'(y) + N(y), \qquad y = b_1, b_1 + 1, \cdots, e_3.$$

Here $h'(y)$, the marked but uncorrupted profile, is given by $h^l(y)$ in (2) or $h^r(y)$ in (3) according as the middle block is shifted left or right. $N(y)$ is an additive white zero-mean Gaussian noise with variance $\operatorname{Var}(N(y)) = \sigma^2$. The centroid



(a)

Fig. 4. Corresponding empirical distributions.

of the three profile blocks are distorted by the additive noise $N(y)$. Let the control blocks have centroids

$$U_1 = c_1 + V_1 \quad \text{and} \quad U_3 = c_3 + V_3$$

where c_i are the uncorrupted centroids and V_i are the random distortion to the centroids due to the additive profile noise $N(y)$. The middle block has been shifted by a size $\epsilon > 0$ so that its centroid becomes

$$U_2 = c_2 + V_2 - \epsilon$$

if it is left shifted, and

$$U_2 = c_2 + V_2 + \epsilon$$

if it is right shifted. Since $N(y)$ is white, the centroid noises $V_i, i = 1, 2, 3,$ are independent.

We have shown in [13] that the centroid noises V_i can be well approximated by zero-mean Gaussian random variables with variance ν_i^2 given by

$$\nu_i^2 = \frac{\sigma^2 w_i}{H_i^2}\left(\delta_i^2 + \left(w_i^2 - 1\right)/12\right) \tag{7}$$

$$H_i = \sum_{b_i}^{e_i} h(y) \tag{8}$$

$$w_i = e_i - b_i + 1 \tag{9}$$

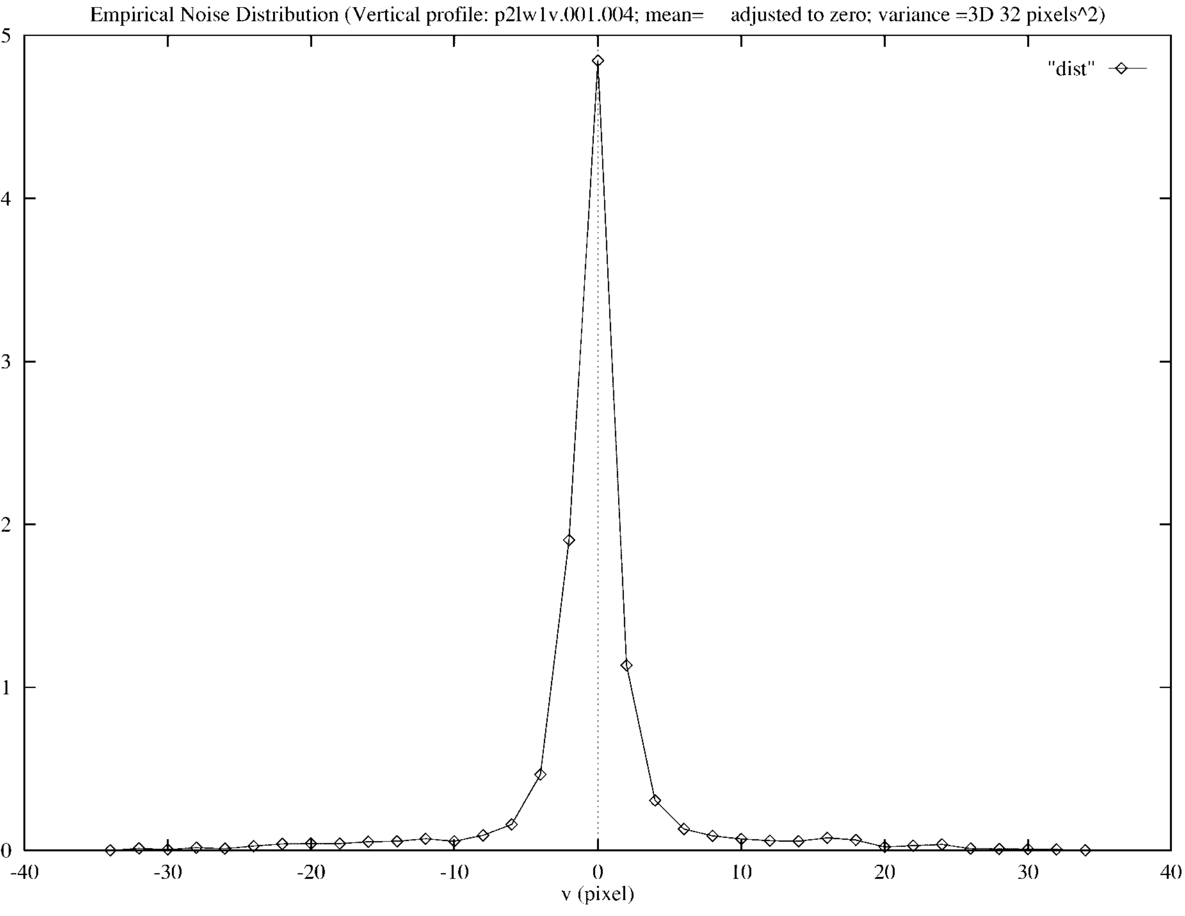
$$\delta_i = c_i - \frac{e_i + b_i}{2}. \tag{10}$$

To eliminate the effect of translation, we base our detection on the relative distance $U_i - U_{i-1}$ of the shifted centroid from the control centroids, instead of on the absolute position of the shifted centroid U_2. We have a classical detection problem in which we have to decide whether the middle centroid has been left or right shifted given the observed values of $U_2 - U_1$ and $U_3 - U_2$. We next derive the maximum-likelihood detection that chooses the direction of the shift that is most likely to have caused the observed $U_2 - U_1$ and $U_3 - U_2$.

It is convenient to use as decision variables the differences

$$\Gamma_l := (U_2 - U_1) - (c_2 - c_1)$$
$$\Gamma_r := (U_3 - U_2) - (c_3 - c_2)$$

of the corrupted centroid separations and the uncorrupted separations. Γ_l is the change in the distance of the middle block from the left control block and Γ_r is that from the right control block. Without noise $\Gamma_l = -\epsilon$ and $\Gamma_r = \epsilon$ if the middle block is left shifted, and $\Gamma_l = \epsilon$ and $\Gamma_r = -\epsilon$ if it is right shifted. Hence it is reasonable to decide that the middle block is left shifted if $\Gamma_l \leq \Gamma_r$, and right shifted otherwise. With noise, according to the following proposition, these changes in the distance of the middle block from the control blocks should be weighted by the noise variances in the centroids of the control blocks before being compared. Note that the decision does not depend on the middle block, except through Γ_l and Γ_r.



(b)

Fig. 4. *(Continued.)* Corresponding empirical distributions.

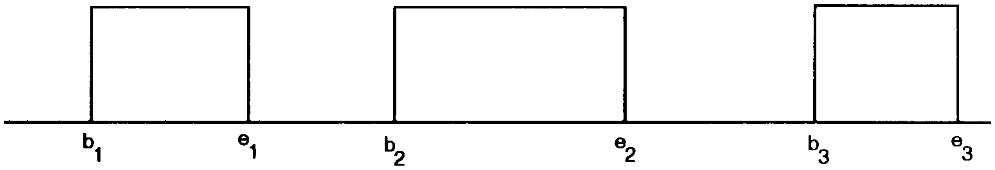


Fig. 5. Profile $h(y)$.

The proofs of the following two propositions can be found in [13].

Proposition 3: The maximum-likelihood decision, when the observed value of (Γ_l, Γ_r) is (γ_l, γ_r), is

$$\begin{array}{ll} \text{left shift} & \text{if } \gamma_l/\nu_1^2 \le \gamma_r/\nu_3^2 \\ \text{right shift} & \text{otherwise} \end{array}$$

where ν_1^2 and ν_3^2 are the centroid noise variances of the left and right control blocks, respectively, given in (7).

Note that the test in the proposition does not require measurement of the profile noise variance σ^2 since it appears in both ν_1^2 and ν_3^2 in (7). Only the three parameters H_i, w_i, δ_i of each uncorrupted control block are necessary.

We evaluate the performance of this decision rule using the average probability of error given by (6).

Proposition 4: The error probability of the maximum likelihood detector in Proposition 3 is

$$P_E = \text{erf}\left(-\epsilon\sqrt{\frac{\nu_1^2+\nu_3^2}{\nu_1^2\nu_3^2+\nu_2^2(\nu_1^2+\nu_3^2)}}\right)$$

where $\text{erf}\,(x) := (2\pi)^{-1/2}\int_{-\infty}^{x} e^{-y^2/2}\,dy$.

IV. Performance Comparisons

In this section we make four performance comparisons based on the probabilities of error derived in the last section. For each detection method we compare line and word-shift detection, and for each type of profile we compare the centroid and correlation detectors.

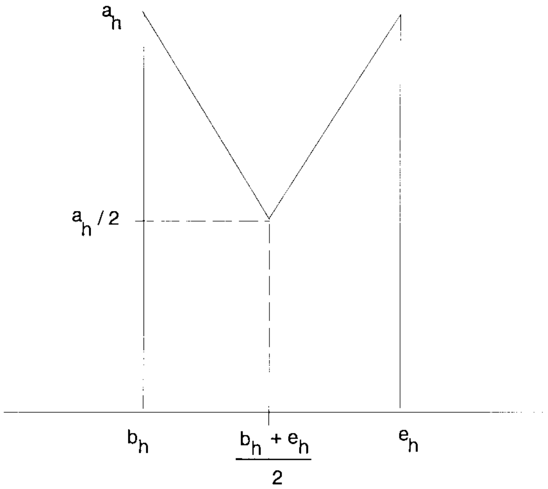


Fig. 6. Model of a horizontal profile.

A. Model and Summary

The error probabilities are functions of the uncorrupted profile $h(y)$ which is determined by the particular document under detection. Centroid detection depends on $h(y)$ mainly through two parameters, its total weight and width. We can characterize the dependence of its error probability on these parameters. This characterization then provides a general but qualitative comparison of the centroid detection of line and word shifts (see Section IV-B). Correlation detection on the other hand depends on the entire $h(y)$. To make concrete comparison we use simple models for horizontal and vertical profiles as explained next.

A horizontal profile block is typically thin and tall with variations much smaller than the bulk of the profile height (see Fig. 1). Hence we model each horizontal profile block by a deterministic block given by

$$h(y) = \begin{cases} \frac{a_h}{w_h}(e_h - y), & y \le \frac{1}{2}(b_h + e_h) \\ \frac{a_h}{w_h}(y - b_h), & y > \frac{1}{2}(b_h + e_h) \end{cases} \tag{11}$$

where $w_h := e_h - b_h$ is the width of the block, as shown in Fig. 6.

We model each block of a vertical profile by a stationary process [19] $h(y)$ such that $h(y), b_v \le y \le e_v$, are uncorrelated, identically and uniformly distributed over $[0, a_v]$. Let $w_v = e_v - b_v$ denote its width.

TABLE I
SUMMARY OF PERFORMANCE COMPARISON

	Expression	Typical value
θ_{cl}	$9a_h\epsilon_h^2/2w_h\sigma_h^2$	28
θ_{ol}	$a_h^2\epsilon_h/\sigma_h^2$	247
θ_{cw}	$2a_v^2\epsilon_v^2/w_v\sigma_v^2$	1
θ_{ow}	$a_v^2w_v/24\sigma_v^2$	301

Using these simple models we compare the probabilities of error for both types of profiles under both methods. The figure of merit is a parameter θ such that the error probability $P_E = \text{erf}\,(-\sqrt{\theta})$ (for horizontal profiles) or $P_E \ge \text{erf}\,(-\sqrt{\theta})$ (for vertical profiles). Let θ_{cl} and θ_{ol} be the figures of merit for centroid and correlation detection, respectively, of line shifts, and let $\overline{\theta}_{cw}$ and $\overline{\theta}_{ow}$ be those for centroid and correlation detection, respectively, of word shifts. Then the subsections in the sequel derive these θ's as functions of the following profile parameters (the typical values, in unit of pixel or 1/300 inch at 300 dpi digitization resolution, are from the examples in Figs. 1 and 4):

Definition		Typical value
a_h	horizontal profile noise height	850
w_h	horizontal profile width	40
ϵ_h	size of line shift	1
σ_h^2	horizontal profile noise variance	2916
a_v	vertical profile height	34
w_v	vertical profile width	200
ϵ_v	size of word shift	2
σ_v^2	vertical profile noise variance	32

The derivation is summarized in Table I and illustrated using example values of profile parameters. From the table the centroid detection of line shifts has a smaller error probability than that of word shifts $(\theta_{cl} > \overline{\theta}_{cw})$, illustrating the qualitative statement of Section IV-B. For correlation detection the comparison is less conclusive as $\text{erf}\,(-\sqrt{\overline{\theta}_{ow}})$ provides only a lower bound on the error probability and the typical values of θ_{ol} and $\overline{\theta}_{ow}$ are close. The table suggests that correlation detection outperforms the centroid detection for both line and word shifts. However correlation detection requires accurate compensation of translation of text introduced by the noise process whereas centroid detection does not, and hence centroid detection seems to perform better in practice on line shifts. This has led to our using centroid detection for line shifts and correlation detection for word shifts in our prototype system.

We present the derivation in the next four subsections.

B. Line Versus Word Shift—Centroid Detection

Our comparison is based on the expression for error probability for centroid detection in Proposition 4

$$P_E = \text{erf}\left(-\epsilon\sqrt{\frac{\nu_1^2 + \nu_3^2}{\nu_1^2\nu_3^2 + \nu_2^2(\nu_1^2 + \nu_3^2)}}\right) \tag{12}$$

where $\text{erf}\,(x) := (2\pi)^{-1/2}\int_{-\infty}^{x} e^{-y^2/2}dy$. In a real document profile the deviation δ_i of the uncorrupted centroid from the center is typically negligible compared with the profile width w_i. Hence, from the expression for ν_i^2 in (7), P_E mainly depends on the total weight H_i and width w_i of the profile over each block i. We will first characterize P_E's dependence on H_i and w_i and then compare line- and word-shift detection.

Suppose the height of profile block $i, i = 1, 2, 3$, is uniformly increased by a factor of $\lambda_i > 0$ so that they become

$$\lambda_i h(y), \qquad y = b_i, \cdots, e_i.$$

The next proposition says that P_E decreases with increasing profile height.

Proposition 5: The probability of error $P_E(\lambda_1, \lambda_2, \lambda_3)$ as a function of the scaling factors $\lambda_i > 0$ to the profile height is decreasing in each of its arguments.

Proof: The total weight $H_i = \Sigma_{b_i}^{e_i} \lambda_i h(y)$ is increased by a factor λ_i, and ν_i^2 is decreased by a factor λ_i^2 according to (7). Hence

$$\chi(u_1, u_2, u_3) = \frac{u_1 + u_3}{u_1 u_3 + u_2(u_1 + u_3)}. \qquad (13)$$

is increased as χ is decreasing in each of its arguments. Since

$$P_E(\lambda_1, \lambda_2, \lambda_3) = \operatorname{erf}\left(-\epsilon\chi(\nu_1^2, \nu_2^2, \nu_3^2)\right)$$

and erf is increasing in its argument, the result follows. □

The next proposition shows that the probability of error increases with profile width.

Proposition 6: Suppose the weight of profile block i is spread over a wider interval in such a way that the width w_i is increased but the total weight H_i and the centroid are unchanged. The probability of error $P_E(w_1, w_2, w_3)$ as a function of profile width w_i is increasing in each of its arguments.

Proof: From (7), ν_i^2 increases with w_i, and hence $\chi(\nu_1^2, \nu_2^2, \nu_3^2)$ defined in (13) decreases. The result follows as in the previous proof. □

Since horizontal profiles are generally taller and narrower than vertical profiles (see Fig. 1), in view of the two propositions, line-shift detection generally has a smaller error probability than word-shift detection using the centroid method, as illustrated in Table I.

C. Line Versus Word Shift—Correlation Detection

From Proposition 2 the probability of error for correlation detector is

$$P_E = \operatorname{erf}\left(-\sqrt{\frac{\sum h^2(y) - \sum h^l(y)h^r(y)}{2\sigma^2}}\right) \qquad (14)$$

where $h^l(y)$ and $h^r(y)$ are computed from the original profile $h(y)$ according to (2) and (3), respectively. This probability depends on the value of $\Sigma\, h^2(y) - \Sigma\, h^l(y)h^r(y)$ which, since h^l and h^r coincide with h over the two control blocks, depends only on the middle block

$$\sum_{b_1}^{e_3} h^2(y) - \sum_{b_1}^{e_3} h^l(y)h^r(y) = \sum_{b_2}^{e_2} h^2(y) - \sum_{b_2+\epsilon}^{e_2-\epsilon} h(y+\epsilon)h(y-\epsilon).$$

Unlike in the centroid detection where the error probability depends on the profile $h(y)$ only through three parameters H_i, w_i, δ_i, here, it depends on the entire $h(y)$ over the middle block. To make concrete comparison we use the model described in Section IV-A.

To simplify notation we assume that both y and $h(y)$ take continuous value. Since we are only concerned with the middle block in this subsection, we drop the subscript from b_2, e_2, etc.

The following proposition describes the dependence of error probabilities on profile height and width (assuming profile width much bigger than size of shift).

Proposition 7: 1) The probability of error in correlation detection of line-shift is $\operatorname{erf}(-\sqrt{\theta_{ol}})$ where

$$\theta_{ol} = \frac{a_h^2 \epsilon_h}{\sigma_h^2} \frac{(w_h - 2\epsilon_h)(w_h + \epsilon_h)}{w_h^2} \sim \frac{a_h^2 \epsilon_h}{\sigma_h^2}. \qquad (15)$$

2) The probability of error in correlation detection of word-shift is lower bounded by $\operatorname{erf}(-\sqrt{\overline{\theta}_{ow}})$ where

$$\overline{\theta}_{ow} = \frac{a_v^2}{24\sigma_v^2}(w_v + 6\epsilon_v) \sim \frac{a_v^2 w_v}{24\sigma_v^2}. \qquad (16)$$

Proof: For the horizontal profile model (11), the probability of error is $\operatorname{erf}(-\sqrt{\theta_{ol}})$ where

$$\begin{aligned}\theta_{ol} &:= \frac{1}{2\sigma_h^2}\left\{\int h^2(y)\,dy - \int h^l(y)h^r(y)\,dy\right\}\\ &= \frac{1}{2\sigma_h^2}\left\{\int_{b_h}^{e_h} h^2(y)\,dy - \int_{b_h+\epsilon_h}^{e_h-\epsilon_h} h(y+\epsilon_h)\right.\\ &\qquad \left.\cdot\, h(y-\epsilon_h)\,dy\right\}\\ &= \frac{a_h^2\epsilon_h}{\sigma_h^2}\frac{(w_h - 2\epsilon_h)(w_h+\epsilon_h)}{w_h^2}\end{aligned}$$

proving the first assertion.

For the vertical profile model the conditional probability of error is $\operatorname{erf}(-\sqrt{\Theta})$, conditioned on the random variable Θ defined as

$$\Theta := \frac{1}{2\sigma_v^2}\left\{\int h^2(y)\,dy - \int h^l(y)\,h^r(y)\,dy\right\}. \qquad (17)$$

Hence the probability of error for correlation detection of word shift is $E[\operatorname{erf}(-\sqrt{\Theta})]$ where the expectation $E[\cdot]$ is taken with respect to Θ.

Define the function

$$\varphi(\theta) := \operatorname{erf}\left(-\sqrt{\theta}\right) = \int_{-\infty}^{-\sqrt{\theta}} e^{-x^2/2}\,dx.$$

Since $\theta \geq 0$ and

$$\frac{d^2\varphi}{d\theta^2} = \frac{1}{4} e^{-(\theta/2)}\theta^{-(3/2)}(\theta + 1) > 0$$

φ is convex. Applying Jensen's inequality, the probability of error is

$$E\operatorname{erf}\left(-\sqrt{\Theta}\right) = E\varphi(\Theta) \geq \varphi(E\Theta) = \operatorname{erf}\left(-\sqrt{E\Theta}\right).$$

Denoting $E\Theta$ by $\overline{\theta}_{ow}$, we have from (17)

$$\overline{\theta}_{ow} = E\Theta = \frac{a_v^2}{24\sigma_v^2}(w_v + 6\epsilon_v)$$

proving the second assertion. □

Hence a higher horizontal profile (larger a_h) increases θ_{ol} and decreases error probability, whereas profile width w_h has little effect. For vertical profile, on the other hand, both peak a_v and width w_v increases $\overline{\theta}_{ow}$ and decreases the error probability bound. In analogy to a communication system, this is because the “signal energy” $E\int h^2(y)\,dy$ is $a_v^2 w_v/3$.

Since $\operatorname{erf}(x) = \int_{-\infty}^{x} e^{-y^2/2}\,dy$ is increasing in x, line shift enjoys a smaller probability of error than word shift if $\theta_{ol} > \overline{\theta}_{ow}$, or if ratio of noise variances in a vertical and horizontal profile exceeds a threshold determined by the peaks and widths of uncorrupted profiles

$$\frac{\sigma_v^2}{\sigma_h^2} > \frac{a_v^2}{a_h^2}\,\frac{w_v+6\epsilon_v}{24\epsilon_h}\,\frac{w_h^2}{(w_h-2\epsilon_h)(w_h+\epsilon_h)} \sim \frac{a_v^2 w_v}{24a_h^2\epsilon_h}.$$

From Table I this is not true for the profile and noise example in Figs. 1 and 4. Indeed one might suspect that if we encode just a single bit per line by exploiting redundancy in multiple word shifts per line, then the error probability for word shifts might be smaller than that for line shifts under correlation detection. While we have not verified this directly, our experimental results so far suggest that whether this might be true depends heavily on the noise on the profiles. For instance, the experiment reported in [11] shows that, depending on the orientation of the photocopying, word-shift detection with coding (using correlation method) can be at least as reliable as, or far more worse than, line-shift detection (using centroid method that performs better or worse than correlation detection of line shifts depending on noise).

D. Centroid Versus Correlation Detection—Line Shift

We assume for simplicity that each of the three blocks in a horizontal profile, when unmarked and uncorrupted, have the same profile in (11). The error probability in correlation detection is $\operatorname{erf}(-\sqrt{\theta_{ol}})$ where, from (15)

$$\theta_{ol} = \frac{a_h^2\epsilon_h}{\sigma_h^2}\,\frac{(w_h-2\epsilon_h)(w_h+\epsilon_h)}{w_h^2} \sim \frac{a_h^2\epsilon_h}{\sigma_h^2}.$$

The centroid noise has the same variance ν_i^2 for all three blocks. Hence

$$\frac{\nu_1^2+\nu_3^2}{\nu_1^2\nu_3^2+\nu_2^2(\nu_1^2+\nu_3^2)} = \frac{2}{3\nu_1^2} = \frac{9a_h^2 w_h}{2\sigma_h^2(w_h^2-1)}.$$

The error probability in centroid detection is $\operatorname{erf}(-\sqrt{\theta_{cl}})$ where [from (12)]

$$\theta_{cl} = \frac{9\epsilon_h^2 a_h^2 w_h}{2\sigma_h^2(w_h^2-1)} \sim \frac{9a_h^2\epsilon_h^2}{2w_h\sigma_h^2}.$$

Hence correlation detector has a smaller probability of error if $\theta_{ol}/\theta_{cl} > 1$, i.e., if

$$\frac{\theta_{ol}}{\theta_{cl}} \sim \frac{2w_h}{9\epsilon_h} > 1$$

which is typically true (but see comment at the end of Section IV-A).

E. Centroid Versus Correlation Detection—Word Shift

From Proposition 7 the probability of error in correlation detection is lower bounded by $\operatorname{erf}(-\sqrt{\overline{\theta}_{ow}})$ where

$$\overline{\theta}_{ow} = \frac{a_v^2}{24\sigma_v^2}(w_v+6\epsilon_v) \sim \frac{a_v^2 w_v}{24\sigma_v^2}.$$

For centroid detection the probability of error depends on

$$\frac{\nu_1^2+\nu_3^2}{\nu_1^2\nu_3^2+\nu_2^2(\nu_1^2+\nu_3^2)}$$

where ν_i^2 is the variance of centroid noise on profile block i. We assume for simplicity that all the three blocks are generated by the same process $h(y), b \le y \le e$, introduced in Section IV-A. From (7)

$$\nu_i^2 = \frac{\sigma_v^2 w_v}{12H^2}\left(w_v^2-1+12\delta^2\right) \ge \frac{\sigma_v^2 w_v\left(w_v^2-1\right)}{12H^2}$$

where the deviation δ of centroid from the center and $H = \int_b^e h(y)\,dy$ are random variables. Here the inequality holds for all sample paths of positive probability. Then

$$\frac{\nu_1^2+\nu_3^2}{\nu_1^2\nu_3^2+\nu_2^2(\nu_1^2+\nu_3^2)} = \frac{2}{3\nu_i^2} \le \frac{8H^2}{\sigma_v^2 w_v(w_v^2-1)}$$

where, again, the inequality holds for all sample paths of positive probability. The probability of error in centroid detection is thus lower bounded by $E\operatorname{erf}(-\sqrt{\Theta})$ where the expectation is taken with respect to the random variable Θ defined as

$$\Theta := \frac{8\epsilon_v^2 H^2}{\sigma_v^2 w_v(w_v^2-1)}.$$

The same argument as in the proof of Proposition 7(2) shows that this probability is lower bounded by $\operatorname{erf}(-\sqrt{\overline{\theta}_{cw}})$ where

$$\overline{\theta}_{cw} := E\Theta = \frac{8\epsilon_v^2 EH^2}{\sigma_v^2 w_v(w_v^2-1)} = \frac{2a_v^2 w_v\epsilon_v^2}{(w_v^2-1)\sigma_v^2} \sim \frac{2a_v^2\epsilon_v^2}{w_v\sigma_v^2}.$$

Hence correlation detector has a smaller error probability bound than centroid detector if $\overline{\theta}_{ow} > \overline{\theta}_{cw}$, i.e., if

$$\frac{\overline{\theta}_{ow}}{\overline{\theta}_{cw}} \sim \frac{w_v^2}{48\epsilon_v^2} > 1.$$

V. Concluding Remarks

A way to discourage illicit redistribution of text documents is to mark each document copy so that the original recipient can be identified from an illicit copy. We have presented two maximum-likelihood detectors to detect document markings from a noisy copy, one based on profile measurement and the other based on centroid measurement. Using their probabilities of error, we have compared their relative performance in detecting line and word shifts. Our analysis suggests that for word shifts, correlation detection outperforms centroid detection; for line shifts, both methods have about the same performance provided certain distortions on line and word profiles can be compensated for.

From experience, translation of the entire text can sometimes be hard to compensate for accurately on a horizontal profile, in which case correlation detector performs poorly. A horizontal profile consists of distinct tall and narrow columns that can be approximated by delta functions situated at their centroids. This suggests using the centroid detection for line shifts. The effect of translation of the entire text is eliminated by making detection decision based on the distance of the shifted centroid relative to its two control centroids.

These results have been used to design and implement a document marking and identification system that is robust against severe distortions in either the vertical or the horizontal direction. Directional distortion seems to be typical on today's copiers. Our system uses a strategy proposed earlier in [12] which takes advantage of the possibility that the vertical and horizontal profiles can be distorted to different degrees. A line is marked both vertically using line shifting and horizontally using word shifting. To detect the marking, the probability of detection error on horizontal and vertical profiles are estimated using control lines and control groups, respectively. Detection is then made using the less noisy profile. The system uses the centroid method to detect line shifts and the correlation method to detect word shifts. It has reliably detected markings using line and word shifts of 1/150 in from photocopies of up to the 10th generation, from facsmile copies, and from document bitmaps that have gone through lossy compression. These experiments have been reported in [11].

REFERENCES

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

[2] D. Boneh and J. Shaw, "Collusion secure fingerprinting for digital data," Princeton Computer Science Department, Tech. Rep. CS-TR-468-94, 1994.

[3] J. Brassil, S. Low, N. Maxemchuk, and L. O'Gorman, "Electronic marking and identification techniques to discourage document copying," in *Proc. Infocom'94*, June 1994, pp. 1278–1287.

[4] ——, "Electronic marking and identification techniques to discourage document copying," *IEEE J. Select. Areas Commun.*, vol. 13, no. 8, pp. 1495–1504, Oct. 1995.

[5] G. Caronni, "Assuring ownership rights for digital images," in *Proc. Reliable IT Systems, VIS'95*.

[6] A. K. Choudhury, N. F. Maxemchuk, S. Paul, and H. Schulzrinne, "Copyright protection for electronic publishing over computer networks," *IEEE Network*, vol. 9, no. 3, pp. 12–21, May/June 1995.

[7] I. Cox, J. Kilian, T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," in *Proc. 1st Int. Workshop Information Hiding*, pp. 183–206.

[8] Special issue on Copyright and Privacy Protection, *IEEE J. Select. Areas Commun.*, this issue.

[9] E. Koch, J. Rindfrey, and J. Zhao, "Copyright protection for multimedia data," in *Proc. Int. Conf. Digital Media and Electronic Publishing*, 1994.

[10] E. Koch and Z. Zhao, "Toward robust and hidden image copyright labeling," in *Proc. 1995 IEEE Workshop on Nonlinear Signal and Image Processing*.

[11] S. H. Low, A. M. Lapone, and N. F. Maxemchuk, "Document identification to discourage illicit copying," in *Proc. Globecom'95*, Singapore, Nov. 1995, 1203–1208.

[12] S. H. Low, N. F. Maxemchuk, J. T. Brassil, and L. O'Gorman, "Document marking and identification using both line and word shifting," in *Proc. Infocom'95*, Boston, MA, Apr. 1995.

[13] S. H. Low, N. F. Maxemchuk, and A. M. Lapone, "Document identification for copyright protection using Centroid detection," *IEEE Trans. Commun.*, vol. 46, pp. 372–383, Mar. 1998.

[14] B. M. Macq and J.-J. Quisquater, "Cryptology for digital TV broadcasting," *Proc. IEEE*, vol. 83, no. 6, pp. 944–957, 1995.

[15] K. Matsui and K. Tanaka, "Video-steganography," *IMA Intellectual Property Project Proc.*, vol. 1, pp. 187–206, 1994.

[16] L. O'Gorman, "Image and document processing techniques for the RightPages Electronic Library System," *Int. Conf. Pattern Recognition (ICPR)*, Sept. 1992, pp. 260–263.

[17] ——, "The document spectrum for structural page layout analysis," *IEEE Trans. Pattern Anal. Machine Intelligence*, vol. 15, Nov. 1993.

[18] L. O'Gorman and R. Kasturi, "Document image analysis," in *IEEE Computer Society Tutorial Text Series*, IEEE, 1994.

[19] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 2nd ed. New York: McGraw Hill, 1984.

[20] J. O. Ruanaidh, W. J. Dowling, and F. M. Boland, "Phase watermarking of digital images," in *IEEE Proc. ICIP96*, Lausanne, Switzerland, pp. 239–242.

[21] M. D. Swanson, B. Zhu, and A. H. Tewfik, "Transparent robust image watermarking," in *IEEE Proc. ICIP96*, Lausanne, Switzerland, pp. 211–214.

[22] K. Tanaka, Y. Nakamura, and K. Matsui, "Embedding secret information into a dithered multi-level image," in *Proc. 1990 IEEE Military Commun. Conf.*, Sept. 1990, pp. 216–220.

[23] H. Van Trees, *Detection, Estimation, and Modulation Theory*, vol. I. New York: Wiley, 1968.

[24] R. G. van Schyndel, A. Z. Tirkel, and C. F. Osborne, "A digital watermark," in *Int. Conf. on Image Processing*, Austin, TX, vol. 2, pp. 86–90, 1994.

Steven H. Low (S'88–M'92), for a photograph and biography, see this issue, p. 451.

Nicholas F. Maxemchuk (F'89), for a photograph and biography, see this issue, p. 451.