Pergamon

Pattern Recognition, Vol. 30, No. 9, pp. 1505–1519, 1997

0031-3203/97 $17.00+.00

PII: S0031-3203(96)00157-4

SKEW DETECTION AND TEXT LINE POSITION DETERMINATION IN DIGITIZED DOCUMENTS[1]

B. GATOS,[†,‡] N. PAPAMARKOS[†,*] and C. CHAMZAS[†]

[†]Electric Circuits Analysis Laboratory, Department of Electrical and Computer Engineering, Democritus University of Thrace, 67100 Xanthi, Greece

[‡]Institute of Informatics and Telecommunications, National Research Center "Democritos", 15310 Athens, Greece

(*Received* 8 *June* 1995; *in revised form* 30 *July* 1996; *received for publication* 15 *October* 1996)

Abstract—This paper proposes a computationally efficient procedure for skew detection and text line position determination in digitized documents, which is based on the cross-correlation between the pixels of vertical lines in a document. The determination of the skew angle in documents is essential in optical character recognition systems. Due to the text skew, each horizontal text line intersects a predefined set of vertical lines at non-horizontal positions. Using only the pixels on these vertical lines we construct a correlation matrix and evaluate the skew angle of the document with high accuracy. In addition, using the same matrix, we compute the positions of text lines in the document. The proposed method is tested on a variety of mixed-type documents and it provides good and accurate results while it requires only a short computational time. We illustrate the effectiveness of the algorithm by presenting four characteristic examples. © 1997 Pattern Recognition Society. Published by Elsevier Science Ltd.

Skew detection   Hough transform   Character recognition   Segmentation

1. INTRODUCTION

Today, most information is saved, used and distributed by electronic means. Scanners can convert documents stored on papers into a format suitable for computers. The ever increasing application of digital document analysis systems promoted the development of digital text processing units in order to obtain the information which appeared on the digital documents. The main purpose of these systems is the transformation of text images into recognized ASCII characters, which is mainly performed with Optical Character Recognition (OCR) systems. An OCR system often consists of a preprocessing stage,[1–3] a document layout understanding and segmentation stage,[4–7] a feature extraction stage[8–14] and a classification stage.[15–17] Many of these stages are facilitated if the document has not been skewed during the scanning process and its text lines are strictly horizontal. Although there are some techniques for character segmentation that can work on skewed documents too, they are ineffective and involve great computational cost.[18,19] It is therefore preferable, in the preprocessing stage, to determine the skew angle of the digitized documents.

There are several methods for skew detection. These methods are based on the Hough transform,[20–22] projection based approaches[23,24] and the Fourier transform.[25] Hough transform methods are the most popular, but they are computationally expensive. To this end, different fast Hough transform based techniques have been proposed,[26] including the creation of the gray scale "burst image"[20] and the use of only a selected square of the document where only bottom pixels of candidate objects are preserved.[21] A special case of Hough transform for skew detection is proposed by Yan.[21] This method is based on the correlation between two vertical lines in a document, which is generated if one vertical line is shifted relative to the other. This technique gives good results but, due to the shift process, it is computationally expensive. In projection-based methods, the projections of the document onto specific directions are first calculated. The skew angle corresponds to a rotation angle where some projection characteristics are satisfied. Ciardiello *et al.*,[23] using horizontal projection, determine that the skew angle corresponds to a rotation for which the mean square deviation of a projection histogram is maximized. The method proposed by Baird in reference (24) belongs in this category, as well. According to this method, for an orientation angle θ, we project the locations of characters onto an "accumulator" line perpendicular to the projection direction. By this procedure, we can construct a function $A(\theta)$, whose global maximum corresponds to the skew angle. This method gives good accurate results but is computationally expensive due to its preprocessing stage and global maximum finding stage. The method proposed by Postl in reference (25) belongs to the Fourier transform approaches. According to this method, the skew angle

* Author to whom correspondence should be addressed. Tel.: +30-541-26478; fax: +30-541-26947; e-mail: papamark@voreas.ee.duth.gr.

[1]This work was partially supported by the Greek GSRT grant PENED.

corresponds to the direction for which the density of the Fourier space becomes the largest. This method requires computation of the Fourier transform of the document and for this reason, it is computationally expensive.

In this paper, we propose a new skew detection method based on the information existing on a set of equidistant vertical lines. We accept that a text document consists mainly of horizontal text lines. We can further choose from all the image pixels those lying on a set of equidistant vertical lines. For a document, these pixels correspond to pixels of text lines. By using only these pixels we construct a correlation matrix between the vertical lines. We need not relate every pixel of a line to all pixels of the other lines, but only to those pixels that lie in a specific region defined by the expected maximum skew. In this way, we reduce the computation time significantly without sacrificing accuracy. Finally, we form the vertical projection of the matrix, and the skew angle corresponds to the projection's global maximum. As we demonstrate in Example 3, the proposed method also works well with documents that, in addition to the usual horizontal text lines, contain images, line draws and tables.

By using the correlation matrix, after evaluation of the skew angle is accomplished, we can also find the positions of the text lines. Defining a line detection function, we form a horizontal projection of the matrix. The local maxima of this projection give the positions of the text lines. The detection of the text lines is essential for many applications such as for segmentation and character extraction procedures.

The innovations introduced by the new method are the following:

- *Efficiency*. Instead of using all the image pixels, we use only those lying on certain vertical lines defined in the image. This results in a drastic decrease of the calculation time for skew detection when our method is compared with the Hough transform and the cross-correlation method of Yan. The basic matrix used for data storage is of much smaller dimension compared to other methods, which results in a faster algorithm implementation and minimum storage requirements.
- *Accuracy*. The new method extracts the document skew with high accuracy. This can be further improved by using more than two vertical lines, in contrast with the Yan method which uses only two vertical lines. The use of more than two vertical lines improves the accuracy, reduces the possibility of a wrong result due to noise and diminishes the possibility of missing a text line of short length. This happens because the skew detection accuracy depends on the distance between the first and the last vertical lines.
- *Robustness*. The results of the proposed method are robust to the presence of graphics in the document which is not true for methods based on the Hough transform.

Experimental results, using two or more vertical lines, illustrate the applicability of the proposed method for skew correction of documents rotated at several angles. Apart from the skew detection, the information given by the pixels on the vertical lines is also used to determine the positions of text lines. Thus, we facilitate and accelerate the character segmentation stage.

2. IMAGE SMOOTHING AND VERTICAL LINE DATA ACQUISITION

We define the binary text image $B(x, y) \in \{0, 1\}$ with the integer x,y taking values in the ranges $1 \leq x \leq X_{\text{win}}$ and $1 \leq y \leq Y_{\text{win}}$, and assuming that text pixels are assigned the value 1 and background pixels the value 0. All the distances in this paper are expressed in units of pixel distance, i.e. the horizontal distances in terms of horizontal pixel distance and the vertical distances in term of vertical pixel distance.

Image smoothing. Before applying the method for skew detection to $B(x, y)$, we pre-process the image using the horizontal run-length smoothing algorithm (RLSA),[27] so that text lines are transformed to thick solid lines. According to it, if the number of background pixels $[B(x, y) = 0]$ lying between two adjacent horizontal text pixels is less than or equal to a certain threshold T, then these background pixels are converted into text pixels $[B(x, y) = 1]$.

The proper value of T depends on the text characteristics and primarily on the character width. Therefore, the proper threshold value is selected according to the user's experience. We found that a suitable value for T is $T = 0.1X_{\text{win}}$. Figure 1 shows the results of the above procedure applied to a document.

Line data acquisition. We define now a set of two or more vertical lines in the document. Embodying the previous preprocessing tool, we define that a pixel belongs to a vertical line if its distance from it is less than or equal to $T/2$.

Then, we define the pixels of every vertical line K, lying at horizontal distance m from the left margin, through the following line smoothing binary function:

$$\text{line}_K(y) = \begin{cases} 1 & \text{if } \sum_{i=m-T/2}^{m-T/2} B(i, y) \neq 0, \\ 0 & \text{otherwise}, \end{cases} \qquad y = 1, \ldots, Y_{\text{win}}. \tag{1}$$

We can say that the function $\text{line}_K(y)$ depicts text pixel existence at the vertical line K after the line smoothing transformation. In contrast with the Hough transform approach where all the image pixels are used, we will use only the pixels belonging to these vertical lines. Thus, we need less memory and we speed up significantly our algorithm.

3. SKEW DETECTION USING TWO VERTICAL LINES

In this section, we will describe the new method using only two vertical lines.

3.1. Selection of pixels for skew detection

A common characteristic of a text document is the repetition of the horizontal text lines along the vertical

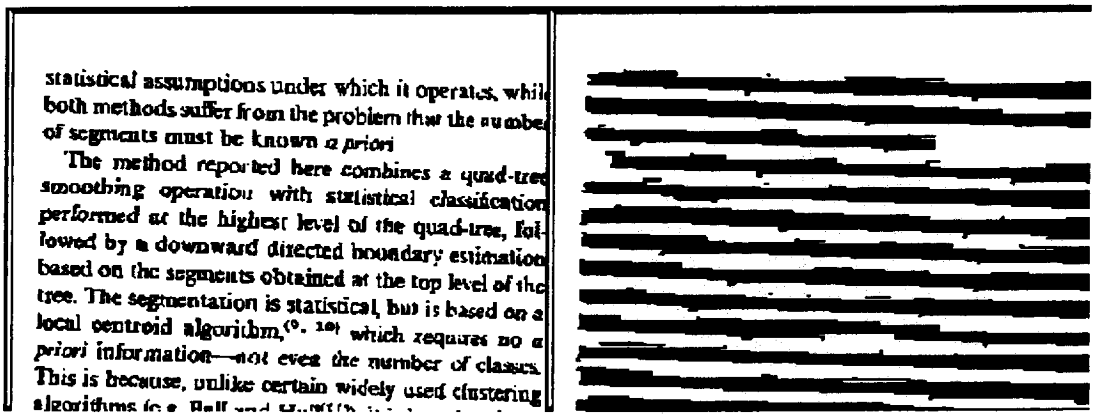


Fig. 1. The image before and after horizontal run-length smoothing.

direction. This is obvious (see Fig. 1) by observing the repetition of the pixel-blocks along the vertical columns. These blocks correspond mainly to horizontal text lines. It is noted that, although in most cases the repetition of text lines is approximately periodical, this is not a prerequirement for our approach. Examination of the blocks between two different vertical lines can give the necessary information for skew detection. We choose two vertical lines d_1 and d_2 (see Fig. 2), at distances D_1 and D_2 from the left margin of the image. Distances D_1 and D_2 are defined so the image is divided into equal parts: $D_1 = \frac{1}{3}X_{\text{win}}$ and $D_2 = \frac{2}{3}X_{\text{win}}$. The skew angle estimation is based on the $2Y_{\text{win}}$ pixels which are on these two lines obtained by equation (1).

3.2. Skew detection from the correlation matrix of two vertical lines

We want to determine a matrix that records all the relative positions of the pixels of the vertical line d_1 to the pixels of the vertical line d_2. We notice that due to the text skew θ, a text line intersects the two vertical lines d_1 and d_2 in two points having vertical distance $l = (D_2 - D_1)\tan\theta$. Making the assumption that a document can be rotated up to $\pm 5^\circ$, i.e. $\theta_{\max}=5^\circ$ due to a scanning misplacement, the vertical distance l must satisfy the constraint:

$$-L < l < L, \quad \text{where } L = (D_2 - D_1)\tan\left(\frac{2\pi\theta_{\max}}{360}\right), \tag{2}$$

where L is an integer, expressed in number of vertical pixels.

For every text pixel of the vertical line d_1 [$\text{line}_1(y_k) = 1$], we search for text pixels at the vertical line d_2 in a region $[-L,L]$ centered at y_k. We store this information in a correlation matrix $C(y_k,\lambda) \in \{0,1\}$ defined as

$$C(y_k,\lambda) = \text{line}_1(y_k)\,\text{line}_2(y_k+\lambda), \quad \text{for } 1 \le y_k \le Y_{\text{win}} \text{ and } -L \le \lambda \le L. \tag{3}$$

Pixels outside the image region are assumed to be 0.

As we can see in Fig. 3, the correlation matrix C has zero elements for y_k=6, 7, 8, 14, 15, 16, 22, 23, 24 and 25. This is because there are no text pixels at line d_1 for these y_k values. We also have $C(1,3)=1$ because there is a text pixel at line d_1 for $y_k=1$ and there is also a text pixel at line d_2 for $y_k=1+3=4$.

If the image skew angle is θ, then the intersection of every text line with the two vertical lines d_1 and d_2 should have a vertical distance $(D_2-D_1)\tan\theta$. So, the correlation matrix C will have maximum accumulation of points along the y-axis for $\lambda = \text{int}[0.5 + (D_2 - D_1)\tan\theta]$ (where int[·] is the integer part of ·). Making a reverse approach of this syllogism, image skew is obtained if we detect the global maximum of the vertical projection of the correlation matrix C.

The vertical projection of the correlation matrix is given from the formula:

$$P(\lambda) = \sum_{k=1}^{Y_{\text{win}}} C(k,\lambda), \quad \forall\lambda \in [-L,L]. \tag{4}$$

According to the above, if the global maximum of $P(\lambda)$ is at $\lambda=\lambda_{\max}$, then the document skew is given by the

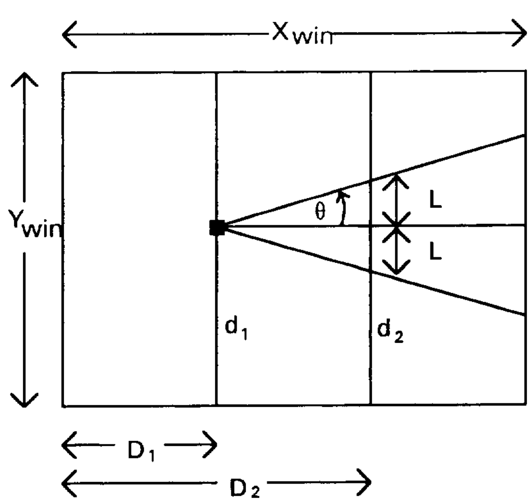


Fig. 2. Image window of size $X_{\text{win}} * Y_{\text{win}}$ and region $[-L,L]$.

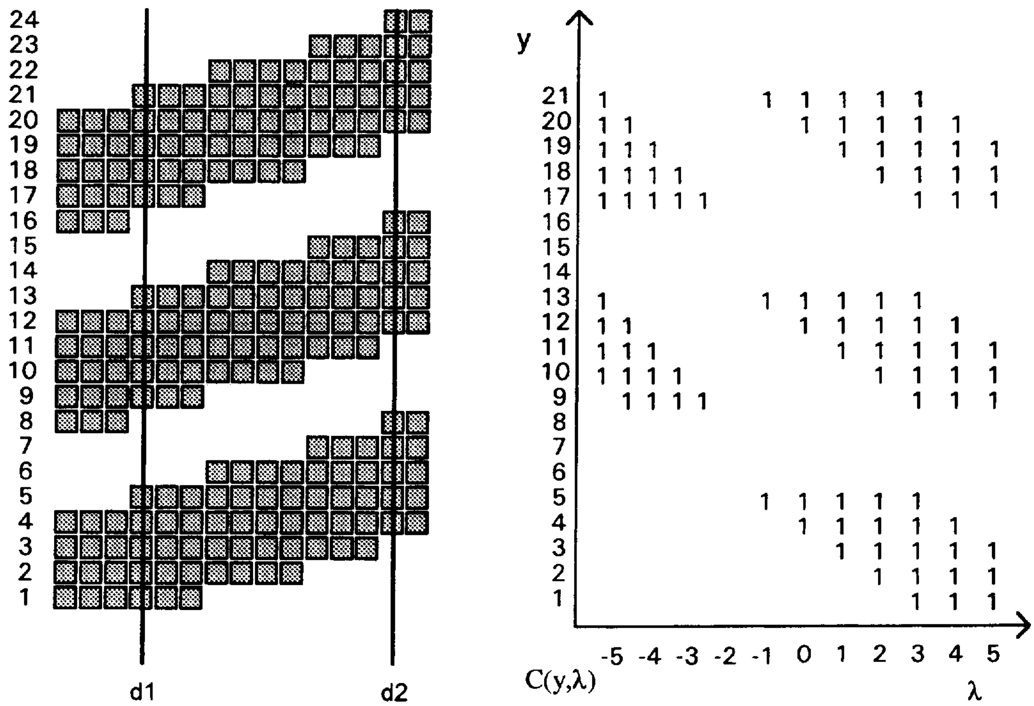


Fig. 3. Correlation matrix of the vertical line d_1 towards d_2.

following relation:

$$\theta = \tan^{-1}\left(\frac{\lambda_{\max}}{D_2 - D_1}\right). \qquad (5)$$

As we can see in Fig. 4, we have a global maximum of the projection for $\lambda=3$, which means that the document skew angle is $\tan^{-1}[3/(D_2-D_1)]$.

3.3. Text line detection using the correlation matrix

With the skew angle of the document determined to be at $\lambda=\lambda_{\max}$, then we can now use the correlation matrix C to compute the position of text lines in the document. The concentration of points around the column of $C(y,\lambda)$ for $\lambda=\lambda_{\max}$ can give us the positions of the text lines. Giving

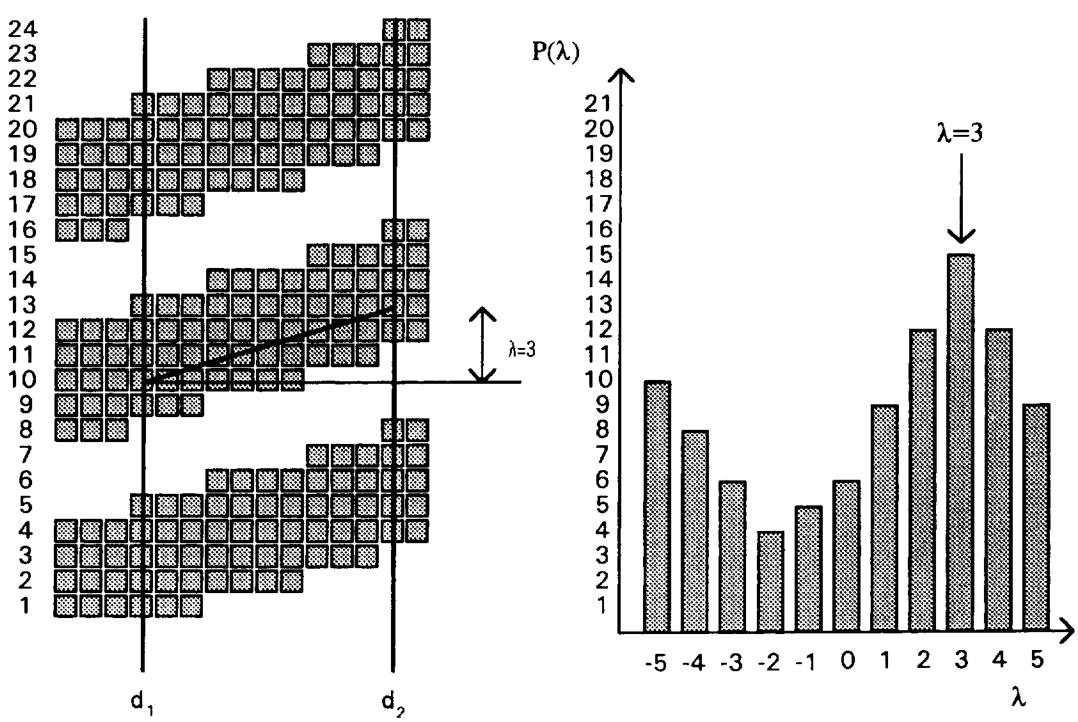


Fig. 4. Skew detection using the correlation matrix of Fig. 3.

suitable weights to the elements of C, starting with 1 for $\lambda=\lambda_{\max}$ and descending by half every time we go away from the point $\lambda=\lambda_{\max}$, we form the line position detection function LI(y):

$$\mathrm{LI}(y) = \sum_{k=-L}^{\lambda_{\max}} \frac{1}{2^{\lambda_{\max}-k}} C(y,k) + \sum_{k=\lambda_{\max}+1}^{L} \frac{1}{2^{k-\lambda_{\max}}} C(y,k), \quad y = 1,\ldots,Y_{\mathrm{win}}. \tag{6}$$

We use the weights $1/2^{|\lambda_{\max}-k|}$ to emphasize the points close to $\lambda_{\max}$. We also, could use other kernel values, i.e. $1/a^{|\lambda_{\max}-k|}$, $a\neq 2$. However, our experimental results indicated that $a=2$ is a suitable choice. Taking $a=2$, we have another advantage that is the fast implementation of equation (6) since $1/2^{\lambda_{\max}-k}$ is a single register shift.

The local maxima of the function LI(y) give us the position of the text lines. That is, if $y_1, y_2, \ldots, y_n$ are the local maxima of the function LI(y), then the text lines lie on the image lines with coordinates:

$$((D_1, y_1-\lambda_{\max}), (D_2, y_1)), ((D_1, y_2-\lambda_{\max}), (D_2, y_2)), \ldots, ((D_1, y_n-\lambda_{\max}), (D_2, y_n)). \tag{7}$$

An example of this procedure is given in Fig. 5, where we have local maxima at y=3, 11 and 19. Using these values and the skew angle, we can determine the positions and the slopes of the text lines. As we can see in Fig. 6, we have detected three lines passing from the y=3, 11 and 19 pixels of line 1 having approximately 3° slope.

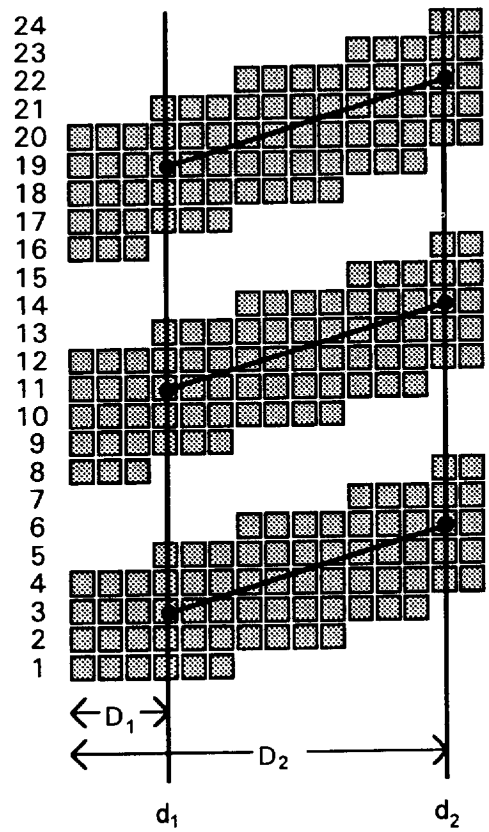


Fig. 6. Lines detected from the three local maxima of Fig. 5 correspond to lines with coordinates: $((D_1,3)-(D_2,6))$, $((D_1,11)-(D_2,14))$, $((D_1,19)-(D_2,22))$.

4. SKEW DETECTION BY USING MULTIPLE VERTICAL LINES

Now, we will extend the skew detection method to more than two vertical lines. Using multiple vertical lines, we can increase the accuracy and robustness of the method.

4.1. Determination of the correlation matrices of multiple vertical lines

In order to increase the method's accuracy and robustness, we use the information of the image pixels that lie on more than two vertical lines. Suppose we have defined M vertical lines d_i, $i = 1,\ldots,M$. Distances D_i from the left margin are defined so that the image is divided into

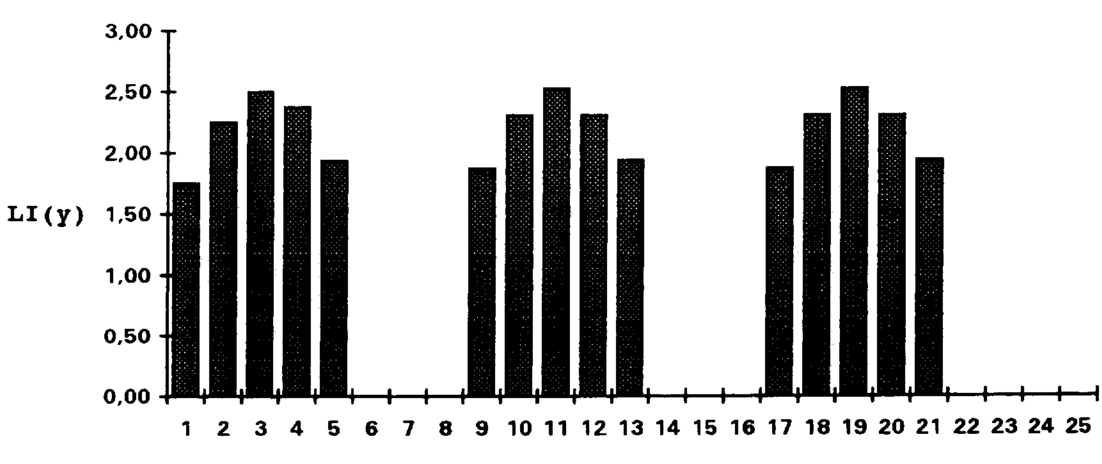


Fig. 5. The line detection function of the correlation matrix of Fig. 3. We detect text lines at local position for y=3, 11 and 19.

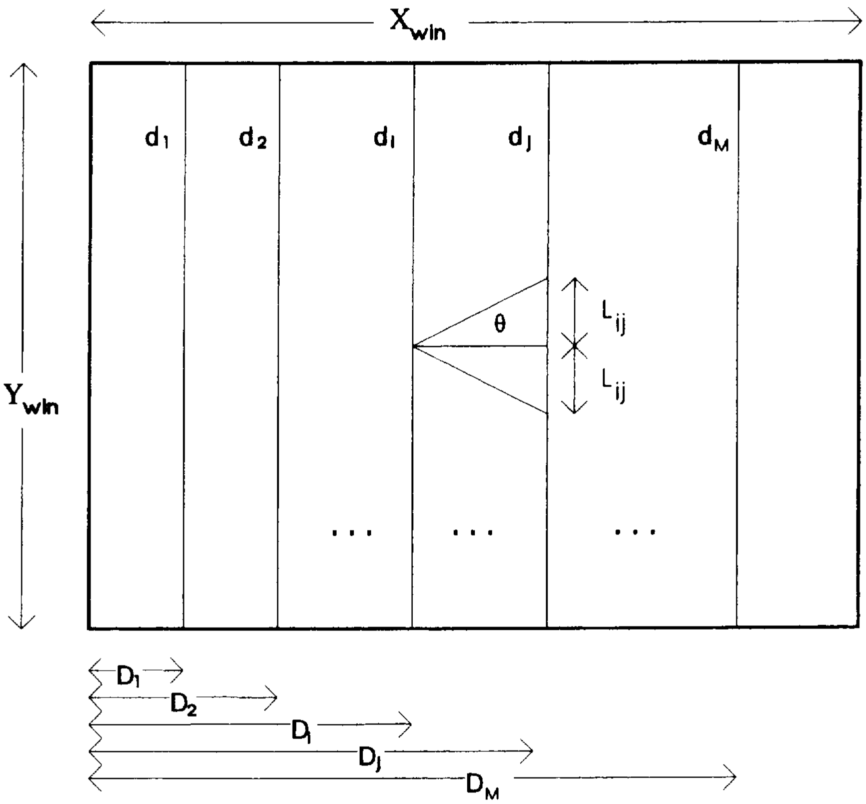


Fig. 7. Image of size $X_{\text{win}} \times Y_{\text{win}}$ and vertical lines at distances $D_1, D_2, \ldots, D_M$.

equal parts: $D_i = (i * X_{\text{win}})/(M + 1)$, $i = 1, \ldots, M$ (see Fig. 7).

For every pair of vertical lines with distances d_i and d_j, we search for all point pairs with vertical distances λ, such as

$$-L_{ij} \leq \lambda \leq L_{ij}, \quad \text{where } L_{ij} = (D_j - D_i)\tan\left(\frac{2\pi\theta_{\max}}{360}\right). \tag{8}$$

Thus for every pixel of a vertical line d_i [$\text{line}_i(y_k)=1$], we examine the pixels at all vertical lines d_j in a region $[-L_{ij}, L_{ij}]$ centered at y_k. Afterwards, these values are stored in the correlation matrices C_{ij}. Clearly, we have $M(M-1)/2$ correlation matrices given by the following relation:

$$C_{ij}(y, \lambda) = \text{line}_i(y)\,\text{line}_j(y + \lambda), \quad 1 \leq y \leq Y_{\text{win}} \text{ and } -L_{ij} \leq \lambda \leq L_{ij}, \tag{9}$$

for $i = 1, 2, \ldots, M-1$ and $j = i+1, i+2, \ldots, M$.

4.2. *Skew detection from the correlation matrices of multiple lines*

In order to calculate the skew angle of the document, we first construct a global correlation matrix CG from all the correlation matrices C_{ij} determined according to equation (9). Then, as in the case of two lines, the skew angle corresponds to the global maximum of the vertical projection of CG. In order to calculate the matrix CG, we first transform the correlation matrices C_{ij} by suitable scaling, considering all the vertical distances of lines d_i, d_j being not at distances D_j, D_i but at distances D_M, D_1, respectively (see Fig. 8). To perform this scaling procedure, every correlation array C_{ij} is transformed into a new matrix CS_{ij} as follows:

$$\text{CS}_{ij}(y, a) = \text{CS}_{ij}(y, a) + C_{ij}(y, \lambda), \quad \forall a \in \left[\left(\lambda - \frac{1}{2}\right)\frac{D_M - D_i}{D_j - D_i}, \left(\lambda + \frac{1}{2}\right)\frac{D_M - D_i}{D_j - D_i}\right] \tag{10}$$

and for every $y = 1, \ldots, Y_{\text{win}}$ and $-L_{ij} \leq \lambda \leq L_{ij}$.

Now, the global correlation matrix CG is equal to

$$\text{CG}(y, \lambda) = \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} \text{CS}_{ij}(y, \lambda). \tag{11}$$

In the similar way, in the two line case [equation (4)], we define the vertical projection of the global correlation matrix CG on the λ-axis by the relation:

$$P(\lambda) = \sum_{k=1}^{Y_{\text{win}}} \text{CG}(k, \lambda), \quad \forall \lambda \in [-L_{1M}, L_{1M}], \tag{12}$$

where L_{1M} is calculated from equation (8).

The document skew angle is calculated from the global projection maximum. If the global maximum of the projection is for $\lambda = \lambda_{\max}$, then the skew angle

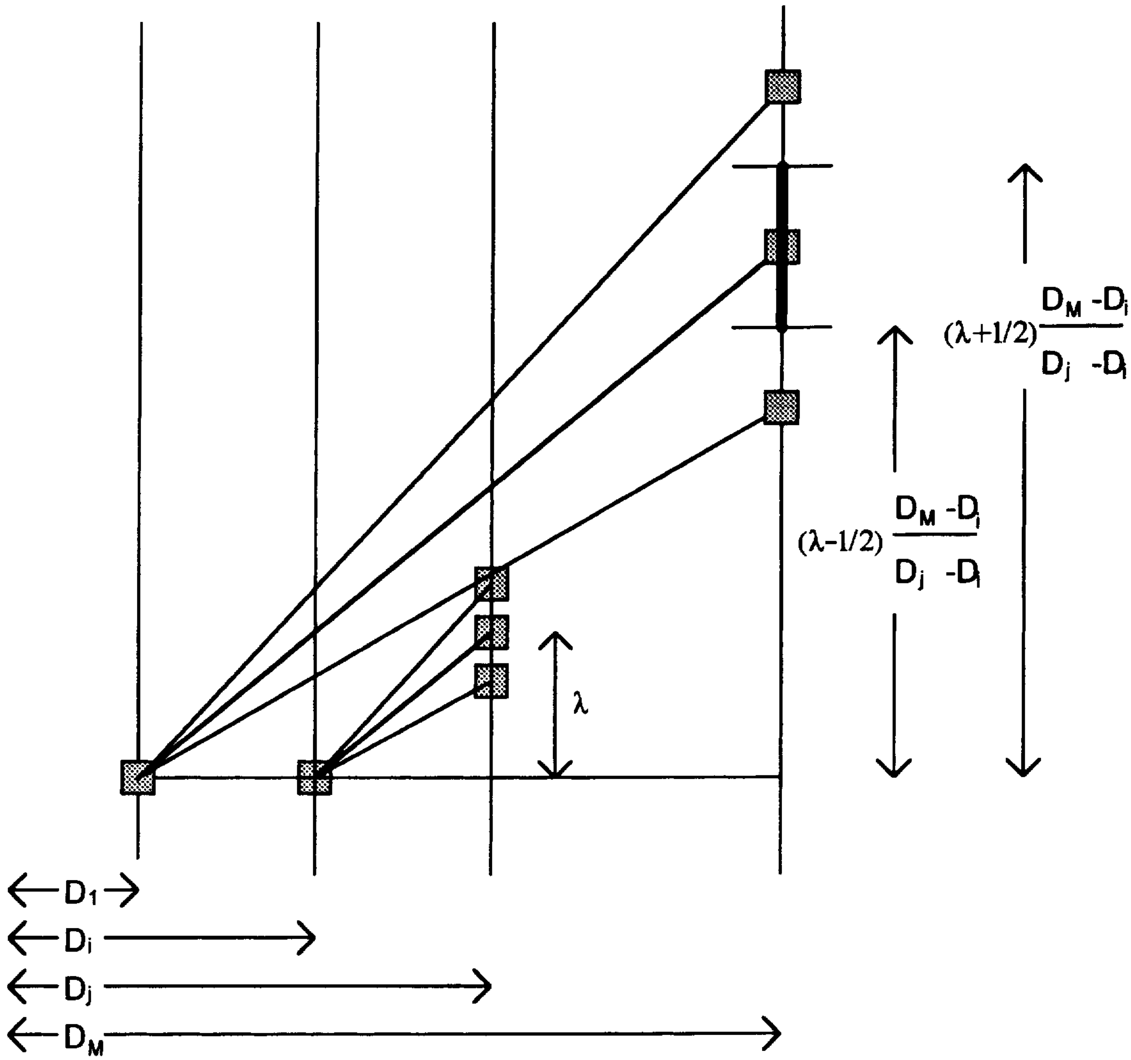


Fig. 8. Normalization of correlation matrices C_{ij}, considering the distances bettwen lines d_i, d_j being not $D_j - D_i$ but $D_M - D_1$.

is equal to

$$\theta = \tan^{-1}\left(\frac{\lambda_{\max}}{D_m - D_1}\right). \tag{13}$$

4.3. Accuracy of the skew detection algorithm

The accuracy of the skew detection algorithm depends on image resolution and on the distance between the first and the last vertical lines. Specifically, the estimated skew angle is $\theta \pm \varepsilon$, where

$$\varepsilon = \frac{1}{2}\tan^{-1}\left(\frac{1}{D_M - D_1}\right). \tag{14}$$

Since $1/(D_M - D_1) \ll 1$, we can approximate ε (in degrees) as

$$\varepsilon \cong \frac{1}{2}\frac{M+1}{M-1}\frac{1}{S \cdot X_{\text{win}}}\frac{360}{2\pi}, \tag{15}$$

where M is the number of vertical lines, S is the scanner's resolution and X_{win} is the horizontal dimension of the document.

As an example, for a digitized image of A4 size (X_{win}=8.5 in.) and with resolution S=300 dpi, we will approximately have:

- For two vertical lines (M=2): $\varepsilon = 0.0337°$.
- For five vertical lines (M=5): $\varepsilon = 0.0169°$.

Table 1. Accuracy of the skew method

No. of vertical lines	Scanner resolution			
	200	300	400	800
2	0.0506	0.0337	0.0253	0.0126
3	0.0337	0.0225	0.0169	0.0084
4	0.0281	0.0187	0.0140	0.0070
5	0.0253	0.0169	0.0126	0.0063
6	0.0236	0.0157	0.0118	0.0059
7	0.0225	0.0150	0.0112	0.0056

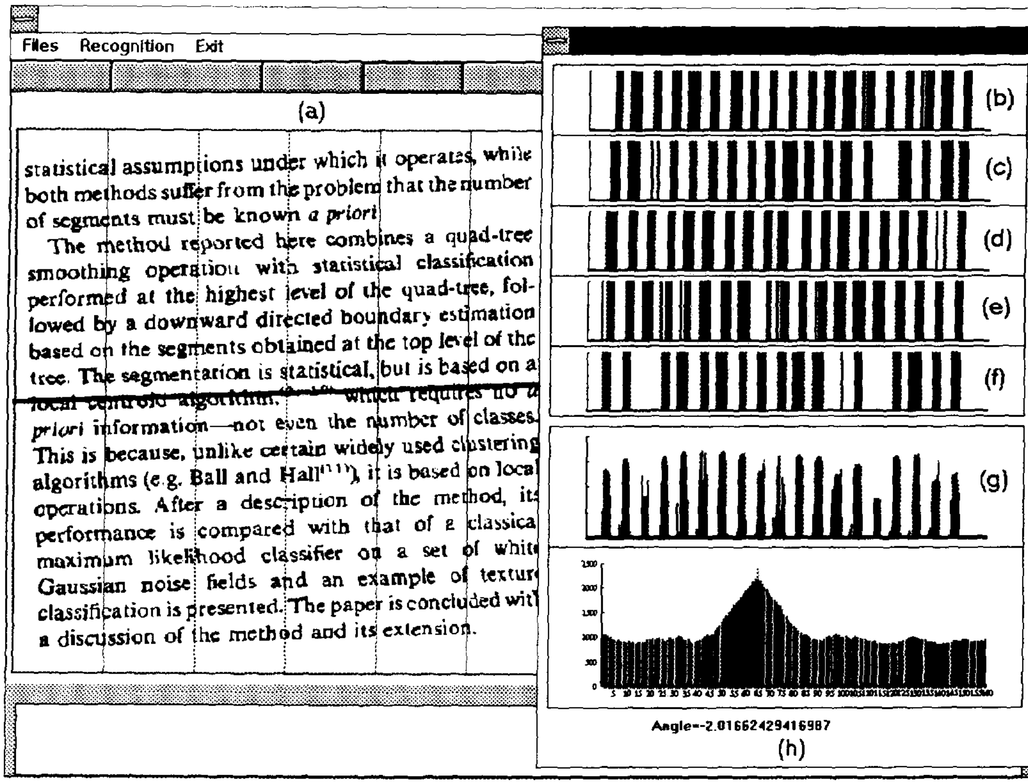


Fig. 9. Application of the algorithm in a simple text with five vertical lines and after line smoothing preprocessing: (a) document; (b)–(f) intersection points of the text with five vertical lines; (g) function LI(y) and (h) $P(\lambda)$.

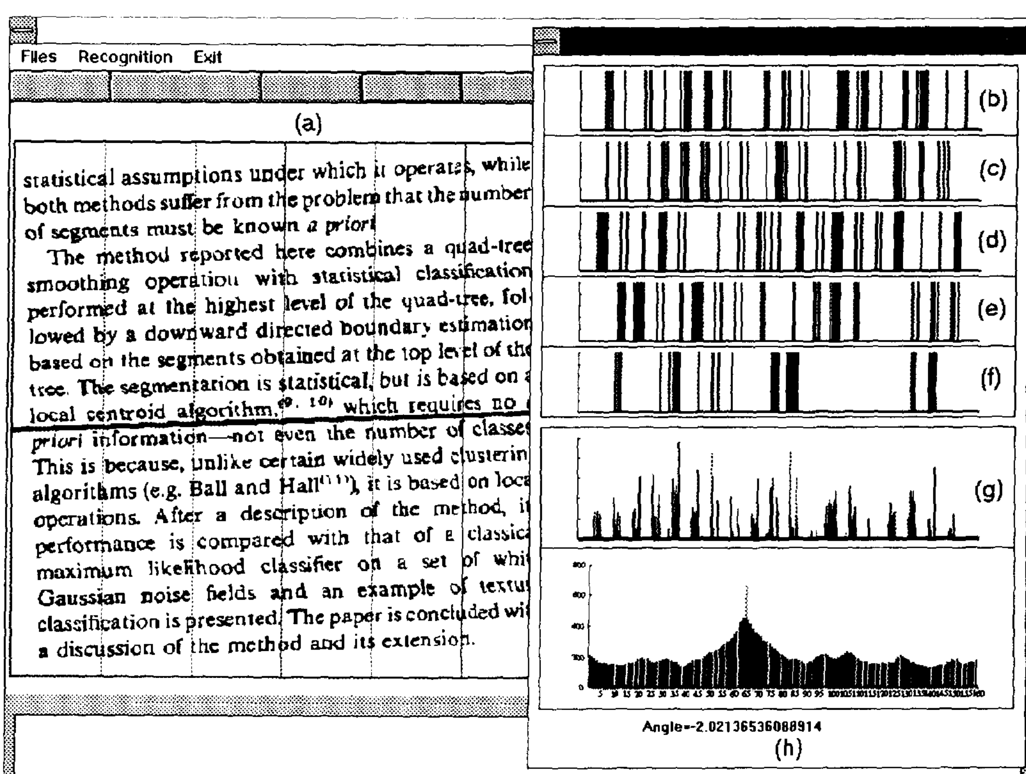


Fig. 10. Application of the algorithm in a simple text with five vertical lines without smoothing preprocessing: (a) document; (b)–(f) intersection points of the text with five vertical lines; (g) function LI(y) and (h) $P(\lambda)$.

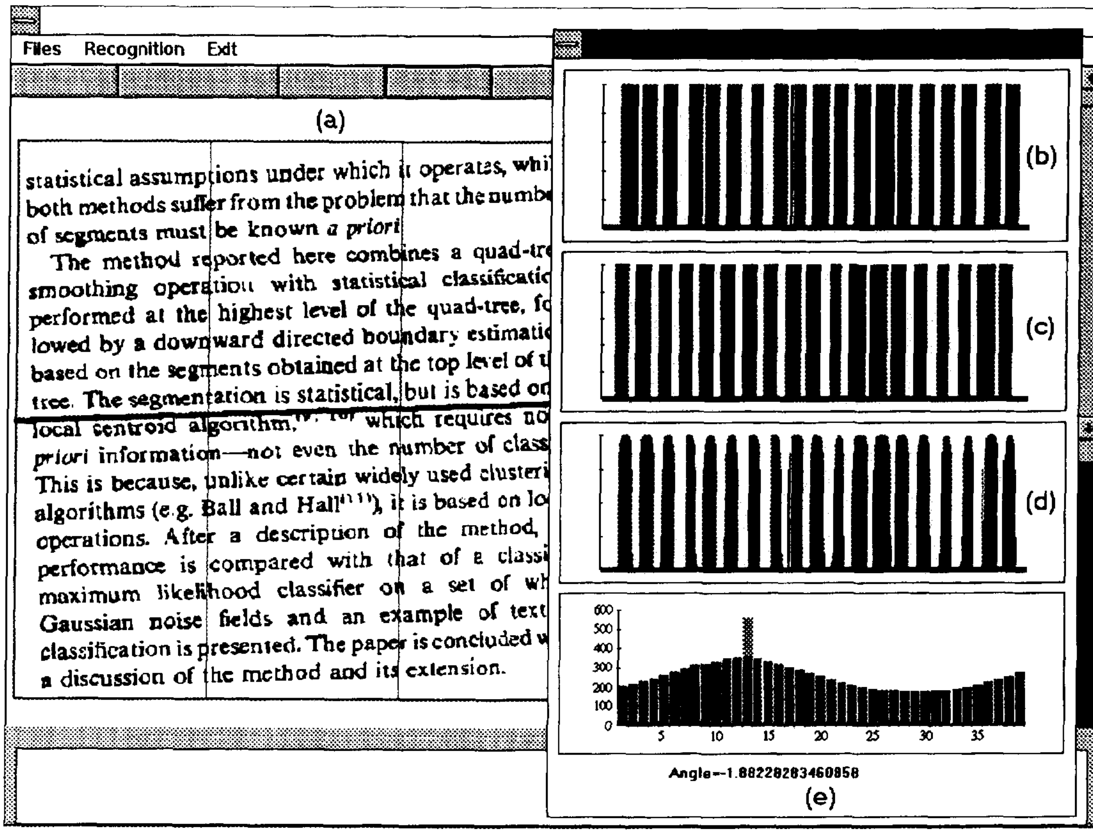


Fig. 11. Example with two vertical lines and smoothing preprocessing: (a) document; (b)–(c) intersection points of the text with two vertical lines; (d) function LI(y); (e) $P(\lambda)$.

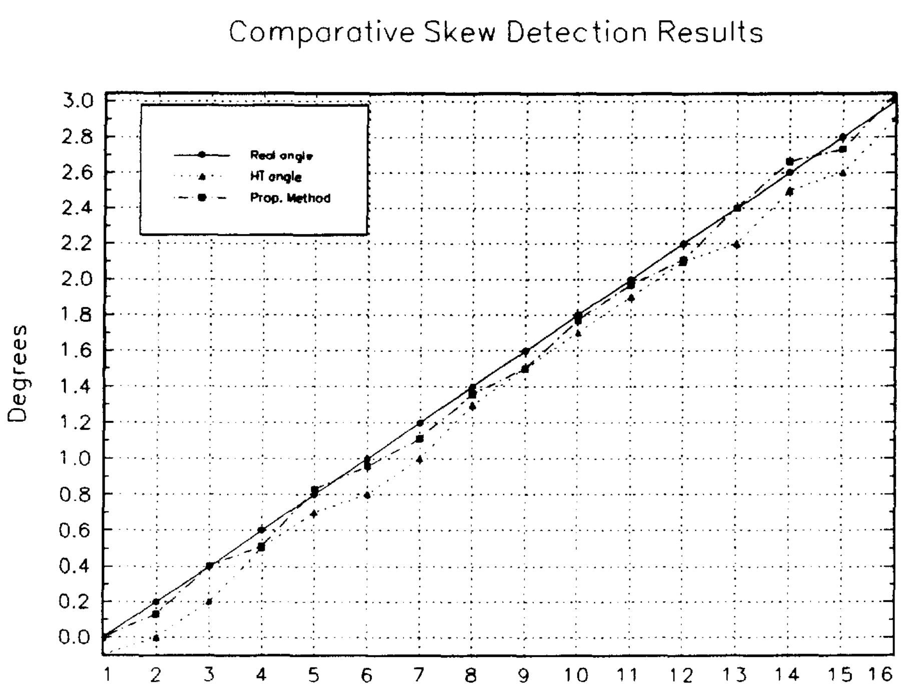


Fig. 12. Comparative skew detection results.

Table 2. Experimental comparative results

Skew angle (degrees)	Hough transform	New method	HT absolute error	Absolute error using the new method
0.0	−0.1	0.0	0.1	0.0
0.2	0	0.13	0.2	0.07
0.4	0.2	0.4	0.2	0.0
0.6	0.5	0.51	0.1	0.09
0.8	0.7	0.83	0.1	0.03
1.0	0.8	0.96	0.2	0.04
1.2	1.0	1.11	0.2	0.09
1.4	1.3	1.36	0.1	0.04
1.6	1.5	1.5	0.1	0.1
1.8	1.7	1.77	0.1	0.03
2.0	1.9	1.97	0.1	0.03
2.2	2.1	2.11	0.1	0.09
2.4	2.2	2.4	0.2	0.0
2.6	2.5	2.66	0.1	0.06
2.8	2.6	2.73	0.2	0.07
3.0	2.9	3.04	0.1	0.04
Maximum error			0.2	0.1
Minimum error			0.1	0.0
Mean value			0.1375	0.04875
Standard deviation			0.0484	0.032763

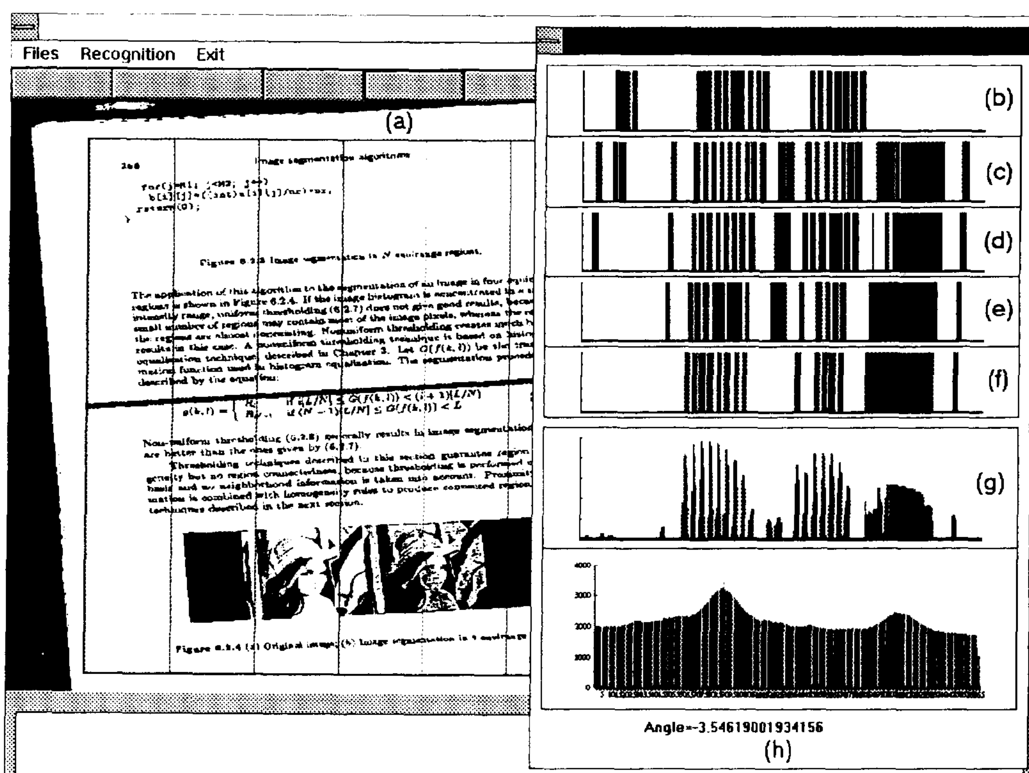


Fig. 13. Application of the algorithm to a document with text and graphic images: (a) document; (b)–(f) intersection points of the text with five vertical lines; (g) function LI(y); (h) $P(\lambda)$.

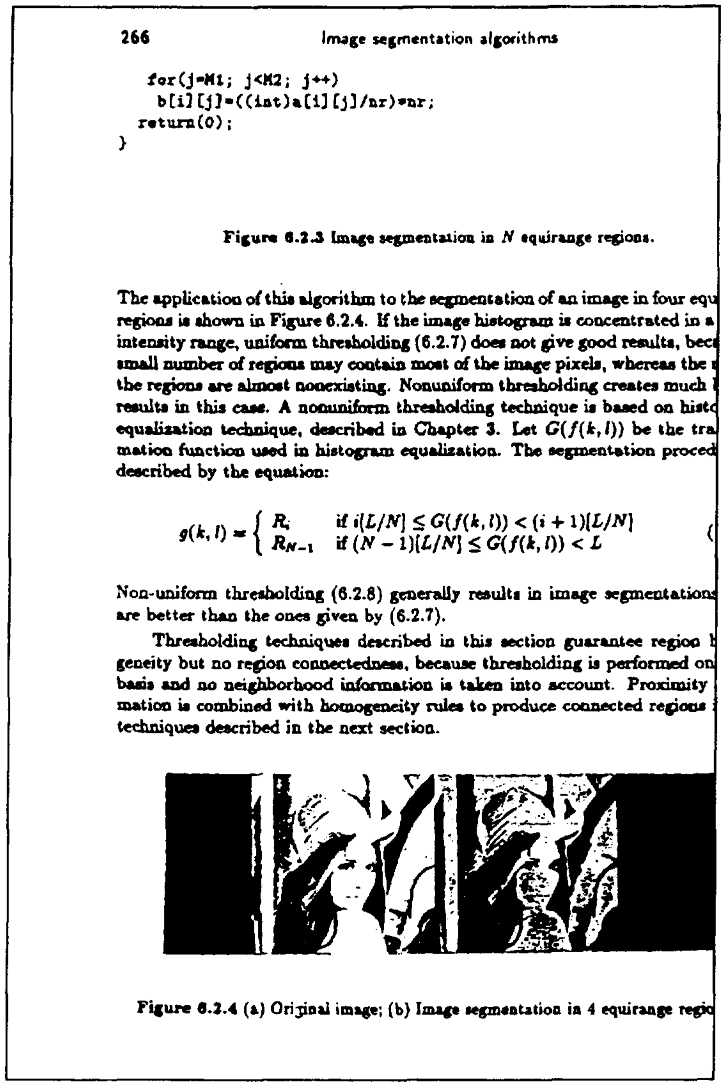
Table 1 shows the accuracy of the skew detection method for some common cases. It is obvious that as the number of vertical lines increases, we improve the accuracy and since we take into account a bigger number of pixels we improve also the robustness.

4.4. Line detection using more than two vertical lines

Having calculated the global correlation matrix CG, as well as the global maximum of the projection for $\lambda=\lambda_{max}$, we can now estimate the positions of the text

Table 3. Experimental comparative results

Skew angle	Image	Hough	Hough+ RLSA	ICC (YAN)	New method
−4°		Result −**4**° Time 334 s Abs. error 0.00°	Result −**4**° Time 470 s Abs. error 0.00°	Result −**3.83**° Time 205 s Abs. error 0.17°	Result −**4.04**° Time 5 s Abs. error 0.04°
−3°		Result −**2.9**° Time 340 s Abs. error 0.10°	Result −**3.1**° Time 470 s Abs. error 0.10°	Result −**3.11**° Time 206 s Abs. error 0.11°	Result −**3.00**° Time 5 s Abs. error 0.00°
−2°		Result −**2.1**° Time 336 s Abs. error 0.10°	Result −**2.2**° Time 476 s Abs. error 0.20°	Result −**1.97**° Time 206 s Abs. error 0.03°	Result −**2.08**° Time 5 s Abs. error 0.08°
−1°		Result −**1.1**° Time 339 s Abs. error 0.10°	Result −**1.2**° Time 471 s Abs. error 0.20°	Result −**1.20**° Time 206 s Abs. error 0.20°	Result −**1.04**° Time 5 s Abs. error 0.04°
0°		Result **0.1**° Time 339 s Abs. error 0.10°	Result **0.1**° Time 473 s Abs. error 0.00°	Result **0**° Time 206 s Abs. error 0.00°	Result −**0.12**° Time 5 s Abs. error 0.12°
1°		Result **0.8**° Time 330 s Abs. error 0.20°	Result **1**° Time 473 s Abs. error 0.00	Result **0.81**° Time 207 s Abs. error 0.19°	Result **0.92**° Time 4 s Abs. error 0.08°
2°		Result **2**° Time 330 s Abs. error 0.00°	Result **2**° Time 476 s Abs. error 0.00°	Result **2.06**° Time 206 s Abs. error 0.06°	Result **1.90**° Time 5 s Abs. error 0.10°
3°		Result **2.9**° Time 327 s Abs. error 0.10°	Result **2.9**° Time 470 s Abs. error 0.10°	Result **2.90**° Time 201 s Abs. error 0.10°	Result **2.93**° Time 4 s Abs. error 0.07°
4°		Result **3.9**° Time 321 s Abs. error 0.10°	Result **3.9**° Time 472 s Abs. error 0.10°	Result **3.75**° Time 202 s Abs. error 0.25°	Result **3.91**° Time 5 s Abs. error 0.09°
Maximum error		0.20°	0.20°	0.25°	0.12°
Minimum error		0.00°	0.00°	0.00°	0.00°
Mean error value		0.088°	0.088°	0.123°	0.068°
Mean required time		332.88 s	472.33 s	205.00 s	4.77 s



266 Image segmentation algorithms

```
  for(j=M1; j<M2; j++)
    b[i][j]=((int)a[i][j]/nr)*nr;
  return(0);
}
```

Figure 6.2.3 Image segmentation in N equirange regions.

The application of this algorithm to the segmentation of an image in four equ regions is shown in Figure 6.2.4. If the image histogram is concentrated in a intensity range, uniform thresholding (6.2.7) does not give good results, bec small number of regions may contain most of the image pixels, whereas the the regions are almost nonexisting. Nonuniform thresholding creates much results in this case. A nonuniform thresholding technique is based on hist equalization technique, described in Chapter 3. Let $G(f(k,l))$ be the tra mation function used in histogram equalization. The segmentation proced described by the equation:

$$g(k,l) = \begin{cases} R_i & \text{if } i[L/N] \leq G(f(k,l)) < (i+1)[L/N] \\ R_{N-1} & \text{if } (N-1)[L/N] \leq G(f(k,l)) < L \end{cases}$$

Non-uniform thresholding (6.2.8) generally results in image segmentation are better than the ones given by (6.2.7).

Thresholding techniques described in this section guarantee region h geneity but no region connectedness, because thresholding is performed on basis and no neighborhood information is taken into account. Proximity mation is combined with homogeneity rules to produce connected regions techniques described in the next section.

Figure 6.2.4 (a) Original image; (b) Image segmentation in 4 equirange regio

Fig. 14. Document of Fig. 13 after skew correction.

lines by calculating the pixel concentration around the column of matrix $CG(y,\lambda_{max})$. The methodology for text line position determination is the same as in the two lines case. We form the text line position function LI and detect its local maxima.

5. EXPERIMENTAL RESULTS

Our method was tested extensively on a variety of document images rotated at various angles. The results were very promising concerning skew angle accuracy and line position determination. In this section, we demonstrate the effectiveness of the method by presenting four characteristic examples implemented in a 486DX/33 MHz computer.

Example 1. In the first example, we apply the method using multiple vertical lines on a document with a skew angle of 2°. The scanning resolution and X_{win} are equal to 96 dpi and 7.398 in., respectively. As we can see in Fig. 9, we use five vertical lines in a pure text area. According to the proposed algorithm, we determine the correlation matrix CG and the line detection function LI(y). In Fig. 9, we can see the intersections of the text with the five vertical lines, the line detection function LI(y) and the projection $P(\lambda)$ of the correlation matrix. The skew angle for this image is calculated to be equal to $-2.0166°$, which is close enough to the real $-2°$.

If we apply the skew detection algorithm without line-smoothing transformation to the same image, then (see Fig. 10) we observe that the skew angle is estimated as $-2.0213°$, which is approximately the same with the one determined in the first case. However, the line detection function takes a non-suitable form and its local maxima cannot be found with great accuracy.

Finally, we apply the proposed skew detection algorithm with two vertical lines, and we estimate now a skew angle at $-1.8822°$ (see Fig. 11).

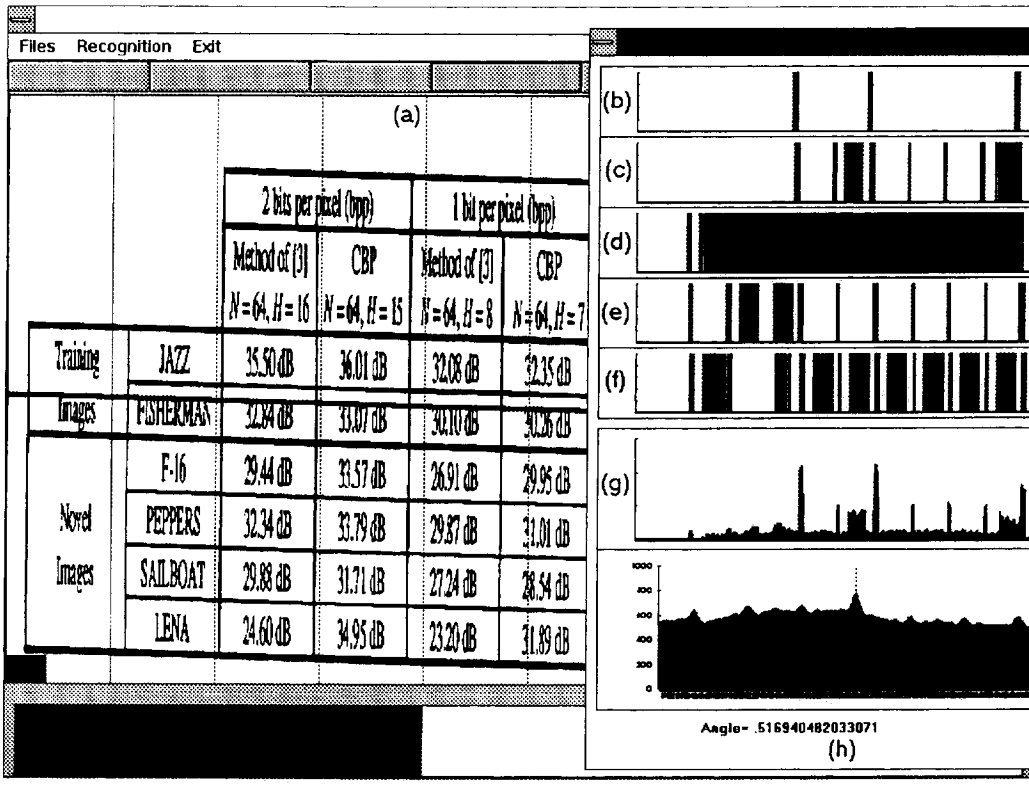


Fig. 15. Application of the algorithm to a document with text into tables: (a) document; (b)–(f) intersection points of the text with five vertical lines; (g) function LI(y); (h) $P(\lambda)$.

Example 2. In order to have comparative results with other skew detection techniques, we applied the Hough transform as well as the proposed method to the same document of Fig. 9 which was skewed from 0° to 3° with a step of 0.2°. Table 2 summarizes the results of this procedure and Fig. 12 shows their graphical representation. As we can see, the maximum absolute error for the Hough transform approach is equal to 0.2°, in contrast with 0.1° for the new method. Additionally, the average computation time was 92 s for the Hough transform and only 7 s for the new method.

Similar results are taken if we compare our method with the use of the Hough transform after smoothing preprocessing and with the cross-correlation method of Yan.[22] The results of this comparison are demonstrated in Table 3 and were obtained from images rotated from −4° to 4° with a step of 1°. The new method gives the smallest maximum error in significantly shorter time.

Example 3. The proposed method works well even when we have mixed texts with graphics or texts into tables. In contrast with the Hough transform method, the new algorithm can be applied to the entire document and not only to pure text areas. We apply the skew detection algorithm, using five vertical lines, to a document of 96 dpi resolution with X_{win}=11.458 in. The real skew angle for this document is −3.5°. As we can see in the result window of Fig. 13, the skew angle was detected properly from the projection of the correlation matrix and equals −3.546°. This slope corresponds to the line appearing in the middle of the image in Fig. 13, while in Fig. 14 we can see the document after skew correction. The Hough transform fails when applied to the same document and results in an unacceptable value of −1.4°.

Example 4. Our skew detection method gives satisfactory results even in cases of documents with tables. In Fig. 15, we have a document with a text-table that has 300 dpi resolution, X_{win}=6.423 and a real skew angle of −0.5°. We apply the skew detection algorithm with five vertical lines. In the result window of Fig. 15, we can see that we have a clear global peak in the projection corresponding to an angle of −0.5169°.

6. CONCLUSION

In this paper, we proposed an efficient algorithm for skew correction and text line position determination of document images. The method uses only a limited number of pixels lying on vertical lines located on specified distances. Based only on these pixels, a correlation matrix is constructed. The skew angle is determined from the global maximum of a projection derived from this matrix. After the skew angle is determined, a text line detection function is also defined and its local maxima give the positions of the text lines.

In summary, the proposed method

- is faster and more accurate than the Hough transform and Yan's cross-correlation approach and it requires less memory because it uses only a small subset of the document pixels;
- is robust since it works equally well with mixed text/graphics documents;
- does not depend on the periodicity of the text lines;
- determines the text line positions;
- its accuracy can be adjusted by increasing the number of vertical lines.

REFERENCES

1. W. H. Abdulla, A. O. M. Saleh and A. H. Morad, A preprocessing algorithm for handwritten character recognition, *Pattern Recognition Lett.* **7**, 13–18 (1988).
2. R. Kasturi, S. T. Bow, W. El-Masri, J. Shah, J. R. Gattiker and U. B. Mokate, A System for interpretation of line drawings, *IEEE Trans. Pattern Analysis Mach. Intell.* **12**(10), 978–992 (October 1990).
3. N. Papamarkos and B. Gatos, A new approach for multilevel threshold selection, *Comput. Vision Graphics Image Process.: Graphical Models and Image Process.* **56**(5), 357–370 (September 1994).
4. F. M. Wahl, K. Y. Wong and R. G. Casey, Block segmentation and text extraction in mixed text/image documents, *Comput. Graphics Image Process.* **20**, 375–390 (1982).
5. K. Y. Wong, R. G. Casey and F. M. Wahl, Document analysis system, *IBM J. Res. Devel.* **26**(6), 647–656 (1982).
6. D. Wang and S. N. Shihari, Classification of newspaper image blocks using texture analysis, *Comput. Vision Graphics Image Process.* **47**, 327–352 (1989).
7. P. Chauvet, J. Lopez-Krahe, E. Taflin and H. Maitre, A system for an intelligent office document analysis, recognition and description, *Signal Process.* **32**, 161–190 (1993).
8. F. W. M. Stentiford, Automatic features design for optical character recognition using an evolutionary search, *IEEE Trans. Pattern Analysis Mach. Intell.* **7**, 349–355 (1985).
9. C. Lettera, M. Majer, L. Maser and C. Paoli, Character recognition in office automation, *Advances in Image Processing and Pattern Recognition*, V. Cappellini and R. Marconi, eds, pp. 191–197. Elsevier, North-Holland, Amsterdam (1986).
10. S. Kahan, T. Pavlidis and H. S. Baird, On the recognition of printed characters of any font and size, *IEEE Trans. Pattern Analysis Mach. Intell.* **9**, 274–287 (1987).
11. E. Persoon and K. S. Fu, Shape discrimination using Fourier descriptors, *IEEE Trans. Systems Man Cybernet.* **7**(2), 170–179 (1977).
12. H. Y. Abdelazim and M. H. Hashish, Automatic reading of bilingual typewritten text, *Proc. VLSI and Microelectronic Applications in Intelligent Peripherals and their Application Networks*, pp. 140–144 (1989).
13. B. Taconet and M. H. Hamza, Recognition and restoration of globally degraded printed characters, *Proc. IASTED Int. Symp. Applied Informatics-AI'89*, pp. 102–105 (1989).
14. G. L. Cash and M. Hatamian, Optical character recognition by the method of moments, *Comput. Vision Graphics Image Process.* **39**, 291–310 (1987).
15. B. Gatos and N. Papamarkos, A novel method for character recognition, *Proc. Forth Int. Conf. on Advances in Communication and Control, COMCON 93*, pp. 493–503, Rhodos, Greece (1993).
16. N. Papamarkos, I. Spiliotis and A. Zoumadakis, Character recognition by signature approximation, *Int. J. Pattern Recognition Artificial Intell.* **8**(5), 1171–1187 (October 1994).
17. B. Jahne, *Digital Image Processing*, 2nd edn. Springer, Berlin (1993).
18. J. P. Bixler, Tracking text in mixed-mode documents, *Proc. ACM Conf. on Document Processing Systems*, pp. 177–185 (1988).
19. L. A. Fletcher and R. Kasturi, A robust algorithm for text string separation from mixed text/graphics images, *IEEE Trans. Pattern Analysis Mach. Intell.* **10**, 910–918 (November 1988).
20. S. C. Hinds, J. L. Fisher and D. P. D'Amato, A document skew detection method using run-length encoding and the Hough transform, in *Proc. Tenth Int. Conf. on Pattern Recognition*, pp. 464–468 (1990).
21. D. S. Le, G. R. Thoma and H. Wechsler, Automated page orientation and skew angle detection for binary document image, *Pattern Recognition* **27**(10), 1325–1344 (1994).
22. H. Yan, Skew correction of document images using interline cross-correlation, *Comput. Vision Graphics Image Process.: Graphical Models and Image Process.* **55**(6), 538–543 (November 1993).
23. G. Ciardiello, G. Scafur, M. T. Degrandi, M. R. Spada and M. P. Roccoteli, An experimental system for office document handling and text recognition, in *Proc. Ninth Conf. on Pattern Recognition*, pp. 739–743 (1988).
24. H. S. Baird, The skew angle of printed documents, in *Proc. SPSE 40th Conf. Symp. Hybrid Imaging Systems*, Rochester, New York, pp. 21–24 (1987).
25. W. Pstl, Detection of linear oblique structures and skew scan in digitized documents, in *Proc. Eighth Int. Conf. on Pattern Recognition*, pp. 687–689 (1986).
26. B. Gatos, S. J. Perantonis and N. Papamarkos, Accelerated Hough transform using rectangular image decomposition, *Electronics Lett.* **32**(8), 730–732 (1996).
27. K. Y. Wong, R. G. Casey and F. M. Wahl, Document analysis system, *IBM J. Res. Devel.* **26**(6), 647–656 (1982).

About the Author — BASILIOS GATOS was born in Athens, Greece, in 1967. He received his Diploma degree in Electrical Engineering from the Democritus University of Thrace in 1992. At present, he is working for his Ph.D. degree at the Electrical and Computer Engineering Department of Democritus University of Thrace, in the field of Optical Character Recognition. Since January 1993, he has been a Postgraduate Scholarship holder of the Institute of Informatics and Telecommunications, National Research center "Demokritos". His current interests include document analysis and image processing. He is a member of the Greek Technical Chamber.

About the Author — NIKOS PAPAMARKOS was born in Alexandroupoli, Greece, in 1956. He received his Diploma degree in Electrical and Mechanical Engineering from the University of Thessaloniki, Thessaloniki, Greece, in 1979 and the Ph.D. degree in Electrical Engineering in 1986, from the Democritus University of Thrace, Greece. From 1987 to 1990, Dr Papamarkos was a Lecturer, from 1990 to 1996, Assistant Professor in the Democritus University of Thrace where he is currently an Associate Professor since 1996. During 1987 and

1992 he also served as a Visiting Research Associate at the Georgia Institute of Technology, U.S.A. His current research interests are in digital signal processing, filter design, optimization algorithms, image processing, pattern recognition and computer vision. Dr Papamarkos is a member of IEEE and a member of the Greek Technical Chamber.

About the Author — CHRISTODOULOS CHAMZAS was born in Komotini, Greece. He received the Diploma degree in Electrical and Mechanical Engineering from the National Technical University of Athens, Athens, Greece, in 1974 and the M.S. and Ph.D. degrees in Electrical Engineering in 1975 and 1979 from the Polytechnic Institute of New York, Farmingdale. From 1979 to 1982, Dr Chamzas was an Assistant Professor with the Department of Electrical Engineering at the Polytechnic Institute of New York. In September 1982 he joined AT&T Bell Laboratories, Holmdel, NJ, where he was a member of the Visual Communication Research Department until 1990. Since September 1990, he has been a member of the Faculty of the Electrical Engineering Department at Democritus University of Thrace, where he is a Director of the Electric Circuits Analysis Laboratory. He has been a major player in the definition, design and implementation of the CCITT/ISO (JBIG, JPEG, etc.), standards for coding, storage and retrieval of images (color and bilevel), an area where he holds six international patents. In 1985–1986, he was a Visiting Professor at the Department of Computer Science, University of Grete, Iraklion, Greece. He has held summer positions in Greece, England, Portugal, as well as at Bell Laboratories. His primary interests are in digital signal processing, image coding multimedia and communications systems. He is currently interested in the implementation of multimedia image database algorithms either with fast software or with VLSI design. Dr Chamzas is a member of the Technical Chamber of Greece, Sigma Xi, an Editor of the IEEE Transactions of Communications and a Senior Member of IEEE.