Extraction of Character/Graphics Images from Document Images with Background Pictures *

Mohamed Kamel †
Aiguo Zhao ‡
† Department of Systems Design Engineering
University of Waterloo
Waterloo, Ontario, Canada, N2L 3G1
‡ Atlantis Scientific Systems Group Inc.
Ottawa, Canada, K2C 0P9

Abstract

In this paper we consider the extraction of binary character/graphics images from grayscale document images with background pictures. This image processing operation is particularly important for document image analysis, automated data entry in geographical information systems, optical character recognition, image transmission and videoconferencing. We briefly review three published extraction techniques namely, the global thresholding technique, the nonlinear adaptive technique, and the local contrast technique, and then present two new extraction techniques: a modified nonlinear adaptive technique and a logical level technique. The performances of the five techniques are compared by testing them on images of typical poor-quality text documents. The results of the experiments and evaluations have shown that the logical level technique is superior to the rest suggesting its suitability for high-speed applications.

1 Introduction

In the application of document image analysis systems to recognize text and/or graphics good quality documents are seldomly encountered. Some common problems which contribute to poor-quality documents are: (1) highlighted characters in various colours. (2) smeared or smudged characters/graphics. (3) nonuniform change in colours due to long term storage. (4) poor writing or printing quality. (5) shadows due to poor lighting conditions when the document image is captured. Although more processing can be made in later phases, these problems are mainly handled in the first phase of a document analysis system. In this phase, a grayscale image is first captured and digitized using a scanner or a video camera, then a binary image is extracted from the original grayscale image using certain extraction technique. These poor-quality documents are also often handled by videoconferencing systems which transmit document images for human recognition. It is sufficient to represent a character/graphics image in a binary image which will be more efficient to transmit and process instead of the original grayscale image. Although various sophisticated extraction techniques have been published [1] - [7], further effort to solve these problems is still necessary.

In section 2, three published techniques for extraction of binary images are briefly reviewed, evaluated and analysed according to their performances on typical poor-quality text document images. In section 3, two new techniques are presented and compared with the published techniques. The last section includes conclusions from this work.

2 Previous Work

A survey on initial work in this field can be found in [8]. Recent techniques originating from various application areas including check image processing [6] [9], OCR [10], blueprint image extraction [1] and videoconferencing systems [7] have been reported. An evaluation of three such techniques viz., the nonlinear function technique [6], the local contrast technique [10], and a second derivative technique which is virtually equivalent to the integrated function technique in [6], has been presented by Palumbo, Swaminathan and Srihari [5].

We focus our attention in this work on the extraction techniques which have relative high-speed performances, and therefore are suitable for high-speed applications. In this section, we briefly review, evaluate and analyse three published techniques which will be compared with our two new techniques later.

We have implemented and applied the five techniques to a number of test images. We present here a representative test image shown in Figure 1. It is a common text document image with highlighted characters and shadows, and with width of 512, height of 480 and gray level range of $[0, 255]$.

2.1 Global Thresholding Technique

This technique is one of the earliest and most simple technique for extracting a binary image from an original grayscale image. A description of it can be

*This work is partially supported by the Natural Science and Engineering Research Council of Canada

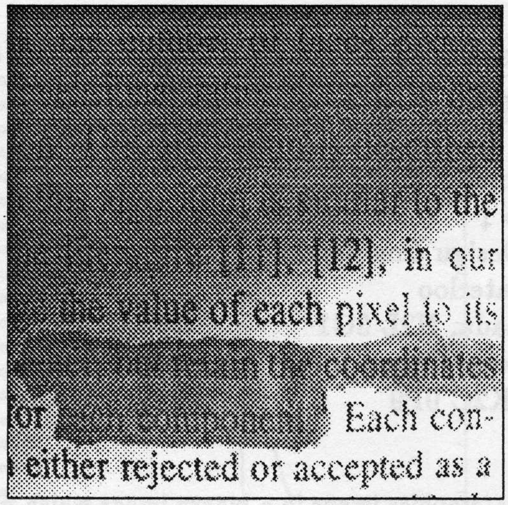

Figure 1: A 512 x 480 x 8 original document image with highlighted characters and shadows

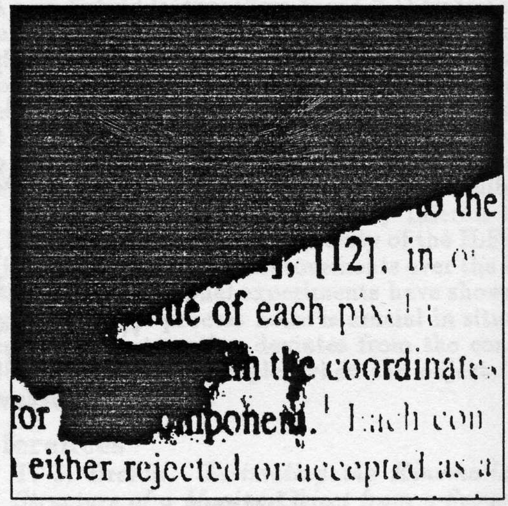

Figure 2: Image extracted using Global Thresholding Technique

found in text books on image processing, see for example [11]. If $f(x,y)$ with $1 \leq x \leq M$, $1 \leq y \leq N$, represents the original grayscale image with width of M, height of N and gray level range of $[0,255]$, and $b(x,y)$ represents the extracted binary character/graphics image with gray level 1 representing character/graphics and 0 representing background, and T be a predetermined constant, then the mathematical description of this technique is

$$b(x,y) = \begin{cases} 1 & \text{if } f(x,y) \leq T \\ 0 & \text{otherwise} \end{cases}$$

This technique has the highest speed due to its most simplest computation, However, it can only handle "ideal" document images of which the gray level distributions of character/graphics pixels and background pixels are well separated away from each other. It can not be applied to the poor-quality documents of which the gray level distributions of character/graphics pixels are often buried within those of background pixels. Figure 2 gives the results of applying this technique to the test image. For the result image, the parameter was manually set to the optimal value based on subjective evaluation. It is obvious that many character/graphics pixels have been lost as background pixels and many background pixels have been misextracted as character/graphics pixels. The reason for this is that this technique does not consider the difference between character/graphics pixels and background pixels and processes every pixel based on its own value without consideration of its neighbouring pixels.

2.2 Nonlinear Adaptive Technique

This technique (as in [6]), is based on comparing the gray level of every pixel with some average of gray levels in a neighborhood, about the pixel, whose size is approximately equal to the character-size. This average is calculated with minimal memory and high-speed using two nonlinear equations. The gray level of the processed pixel is adjusted by a bias function and is then compared with the calculated average. The mathematical description of this technique is as follows:

$$f_H(x,y) = f_H(x-1,y) + W_H[f_H(x-1,y) - f(x,y)]$$

$$f_V(x,y) = f_V(x,y-1) + W_V[f_V(x,y-1) - f_H(x,y)]$$

$$b(x,y) = \begin{cases} 1 & \text{if } Z_b[f(x-L,y)] > f_V(x,y) \\ 0 & \text{otherwise} \end{cases}$$

Where, W_H, W_V are two predetermined update functions. Either W_H or W_V has a value of zero only when its argument has a value of zero, and otherwise has a value between zero and the value of its argument; Z_b is a predetermined bias function; L is a "look ahead" factor; $f_H(x,y)$ is the horizontal average calculated from the gray levels of pixels at all the points (x', y) such that $1 \leq x' \leq x$; $f_V(x,y)$ is the local average calculated from the gray levels of pixels at all the points (x', y') such that $1 \leq x' \leq x$ and $1 \leq y' \leq y$.

The important advantage of this technique is its high speed. However, from its result image, shown in Figure 3, we can see that some pseudo shadows were created in the background and that the left-top part of the dark areas in the background, such as the highlight dark area in the text was also extracted.

2.3 Local Contrast Technique

This technique was developed by Giuliano, Paitra and Stringa [10]. It is virtually a window operator. Every pixel in the output image is calculated using the following operation on a 3x3x5 pixels window.

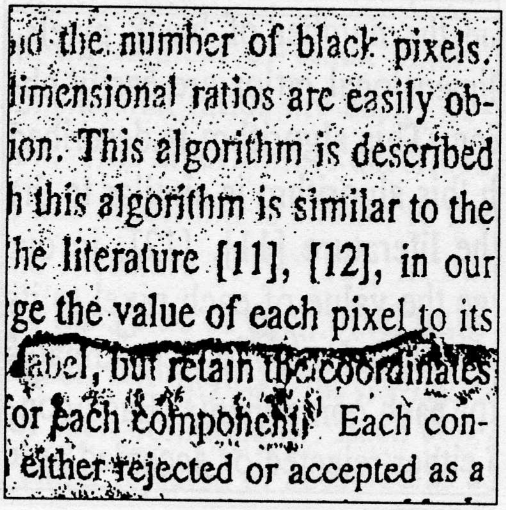

Figure 3: Image extracted using Nonlinear Adaptive Technique

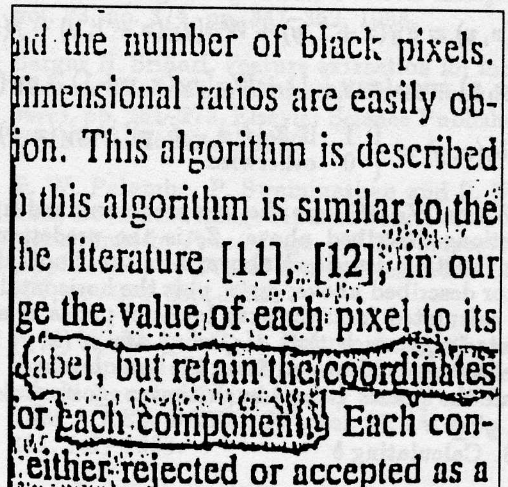

Figure 4: Image extracted using Local Contrast Technique

begin
if ($(f(x,y) < T_1)$ then $b(x,y) = 1$;
else
begin
$a_1 =$ Average of 9 pixels in area A_1;
$A_{2t} = \{ (x,y) \mid (x,y)$ in A_2 and $f(x,y) > T_2 \}$;
$a_2 =$ Average of pixels in area A_{2t};
if $(((T_3 \times a_2) + T_5) > (T_4 \times a_1))$ then $b(x,y) = 1$;
else $b(x,y) = 0$;
end;
end.

Where T_1, a predetermined parameter, is used to extract all the pixels whose gray levels are less than it as character/graphics pixels similar to the parameter in global thresholding technique); T_2, a predetermined parameter, is used to detect all the pixels in A_2 with gray levels not greater than it as unconsidered pixels; T_3, T_4, T_5, three predetermined parameters, are used for comparing a_1 with a_2.

Figure 4 shows the result image of this technique. Compared with the result image of other two published techniques, this image has the best performance with respect to subjective evaluation. However it has two disadvantages. One is its low speed (refer to Table 1), The other is that this technique also extracts the edges of large dark areas in the background, such as the highlight dark area.

3 Two New Techniques

3.1 Modified Nonlinear Adaptive Technique

The major disadvantage of the nonlinear adaptive technique is that it creats some black shadows in the background and extracts the left-top part edges of the large dark areas. Figure 3 illustrates this disadvantage by the black shadows in the check image and the left-top part edge of the highlight dark area in the text document image. One reason for this disadvantage is that the local average $f_V(x,y)$ is calculated based on the pixels in left-top direction of the processed pixel only. Therefore one improvement can be made as follows. After extracting a binary image using the nonlinear adaptive technique, another binary image is also extracted using the same update functions and bias function but different local average which is calculated based on the pixels in the right-bottom direction of the processed pixel. Then the character/graphics image is obtained by performing an *AND* operation on the two binary images.

The mathematical description of this method is as follows:

1. Calculating b_f, the 1st binary image

$$f_H(x,y) = f_H(x-1,y) + W_H[f_H(x-1,y) - f(x,y)]$$

$$f_V(x,y) = f_V(x,y-1) + W_V[f_V(x,y-1) - f_H(x,y)]$$

$$b_f(x,y) = \begin{cases} 1 & \text{if } Z_b[f(x-L,y)] > f_V(x,y) \\ 0 & \text{otherwise} \end{cases}$$

Where W_H, W_V are two predetermined update functions. Either W_H or W_V has a value of zero only when its argument has a value of zero, and otherwise has a value between zero and the value of its argument; Z_b is a predetermined bias function; L is a "look ahead" factor; $f_H(x,y)$ is the horizontal average calculated from the gray levels of pixels at all the points (x', y) such that $1 \le x' \le x$; $f_V(x,y)$ is the local average calculated from the gray levels of pixels at all the points (x', y') such that $1 \le x' \le x$ and $1 \le y' \le y$.

2. Calculating b_g, the 2nd binary image

$$g_H(x,y) = g_H(x+1,y) + W_H[g_H(x+1,y) - f(x,y)]$$

$$g_V(x,y) = g_V(x,y+1) + W_V[g_V(x,y+1) - g_H(x,y)]$$

$$b_g(x,y) = \begin{cases} 1 & \text{if } Z_b[f(x-L,y)] > g_V(x,y) \\ 0 & \text{otherwise} \end{cases}$$

Where W_H, W_V are the two predetermined update functions described above; Z_b is the predetermined bias function described above; L is the "look ahead" factor described above; $g_H(x,y)$ is the horizontal average calculated from the gray levels of pixels at all the points (x',y) such that $x \le x' \le M$; $g_V(x,y)$ is the local average calculated from the gray levels of pixels at all the points (x',y') such that $x \le x' \le M$ and $y \le y' \le N$.

3. Calculating b

$$b(x,y) = \begin{cases} 1 & \text{if } b_f(x,y) = 1 \text{ and } b_g(x,y) = 1 \\ 0 & \text{otherwise} \end{cases}$$

Although both images b_f and b_g contain the character/graphics, some shadows and the edges of large dark areas, their "AND" image b almost contains the character/graphics only, and therefore avoids the major disadvantage of the nonlinear adaptive technique. This improvement is achieved based on two points. On one hand, almost every character/graphics pixel in b_f is also a character/graphics pixel in b_g so that it is kept in the final result image b; on the other hand, most of the shadow and large dark area edge pixels in b_f are not extracted again in b_g so that they are removed in the final result image b. Corresponding to Figure 3 which show the 1st image (b_f), Figure 5 shows the final result image (b) which reveals this improvement. However, this technique has a speed which is approximately one half of the speed of the nonlinear adaptive technique. and has only partially enforced the stroke width restriction. The result images of this technique look a little noisy.

Table 1: Comparison Results of the Five Techniques

Technique	*Subjective Evaluation*	*Width Restr. Enforcement*	*CPU* Sec.
Global Threshold	*worst*	*not*	*1.1*
Nonlinear Adaptive	*pseudo shadows unwanted edges*	*partly*	*6.1*
Local Contrast	*unwanted edges edge*	*partly*	*103.0*
Modified Nonlinear	*a little noise*	*partly*	*10.1*
Logical Level	*best*	*fully*	*7.7*

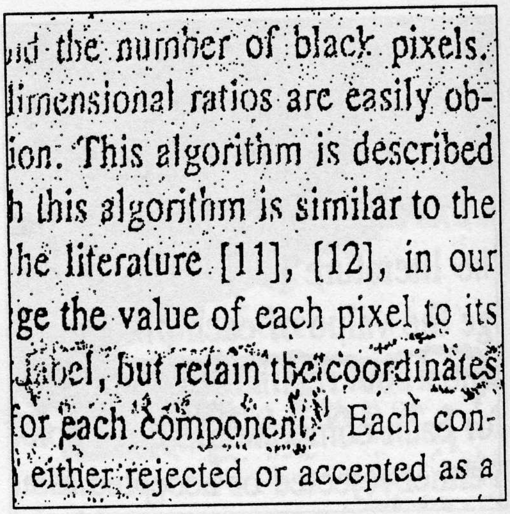

Figure 5: Image extracted using Modified Nonlinear Adaptive Technique

3.2 Logical Level Technique

The nonlinear adaptive technique and the local contrast technique have a common disadvantage that they extract the edges of large dark areas. The reason for this is that they have only partially enforced the stroke width restriction by a comparison of two values. One value is the gray level of the processed pixel or some local average in a small neighborhood about the processed pixel. The other value is some local average in a bigger neighborhood (usually the approximate character-size) about the processed pixel.

From these observations, an outline of a better extraction technique can be to compare the gray level of the processed pixel (if the processed image is noise-free) or its smoothed gray level (if the processed image is noisy) with some local averages in the neighborhoods about a few other neighbouring pixels. Thus more than one comparison is made and the results can be considered as "derivatives". Therefore the labeling, detecting and extracting using the "derivatives", the logical bound on the ordered sequences and the stroke width range can be adopted. Since local averages are not sensitive to noise, these "derivatives" should not be sensitive to noise.

Based on this idea, we developed a new technique called logical level technique [12]. This technique predetermines the stroke width range as $[0, W]$. It processes every pixel by simultaneously comparing its gray level or its smoothed gray level with four local averages in the $(2W-1) \times (2W-1)$-size neighborhoods. For more details see [12].

The major advantages of this technique are its low noise sensitivity, full enforcement of the stroke width and high speed. This technique is not sensitive to noise since it uses local averages as basic computing units. Since few of the large dark area pixels sat-

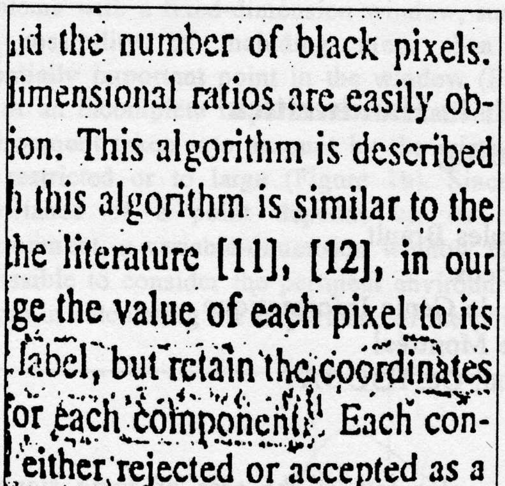

Figure 6: Image extracted using Logical Level Technique

isfy the logical restriction, large dark areas are removed with their edges together. From the result image shown in Figure 6, we can see that this technique has the best performance in terms of subjective evaluation. Unlike the result images of the nonlinear adaptive technique and the local contrast technique, its result images do not contain the unwanted edges of large dark areas in the background, therefore this technique has fully enforced the stroke width restriction. This technique has very high speed which is almost equal to that of the nonlinear adaptive technique (refer to Table 1).

4 Conclusions

In this paper, three published extraction techniques are reviewed, and two new techniques are presented. Comprehensive evaluation and analysis of these five techniques according to their performances on images of typical poor-quality text documents have shown that the new techniques are superior to the existing techniques with respect to quality of extraction and speed of processing.

References

[1] Heng H. Chang, Image thresholding for blueprint image extraction, *Conference Proceedings: Vision '87*, pp. 5-1–5-18, Detroit, Michigan, June 8–11, 1987.

[2] Rangachar Kasturi and Juan Alemany, Information extraction from images of paper-based maps, *IEEE Transactions on Software Engineering*, vol. 14, no. 5, pp. 671–675, May 1988.

[3] A. C. Downton and C. G. Leedham, Preprocessing and presorting of envelope images for automatic sorting using OCR, *Pattern Recognition*, vol. 23, no. 3/4, pp. 347–362, 1988.

[4] Sargur N. Srihari, Feature extraction for locating address blocks on mail pieces, *From Pixels to Features*, pp. 261–273, Elsevier Science Publishers B. V. (North-Holland), 1989.

[5] P. W. Palumbo, P. Swaminathan and S. N. Srihari, Document image binarization: evaluation of algorithms, in *Proceedings, SPIE Symposium on Applications of Digital Image Processing IX*, vol. 697, pp. 278–285, 1986.

[6] J. M. White and G. D. Rohrer, Image thresholding for optical character recognition and other applications requiring character image extraction, *IBM J. RES. DEVELOP.*, vol. 27, no. 4, pp. 400–411, July 1983.

[7] Joan L. Mitchell, William B. Pennebaker, Dimitris Anastassiou and Keith S. Pennington, Graphics image coding for freeze-frame videoconferencing, *IEEE Transactions on Communications*, vol. 37, no. 5, pp. 515–522, May 1989.

[8] Morton Nadler, Survey: document segmentation and coding techniques, *Comput. Vision Graphics and Image Process.* **28**, 1984, 240–262.

[9] Yasuhiko Yasuda, Michael Dubois and Thomas S. Huang, Data compression for check processing machines, in *Proceedings of IEEE*, vol. 68, no. 7, pp. 874–885, July 1980.

[10] E. Giuliano, O. Paitra and L. Stringa, Electronic character reading system, *U. S. Patent*, no. 4,047,15, dated Sept. 6, 1977.

[11] Rafel C. Gonzalez and Paul Wintz, *Digital Image Processing*, Addison-Wesley, 2nd edition, 1987.

[12] Mohamed Kamel and Aiguo Zhao, Extraction of binary character/graphics images from grayscale document images, accepted in *Comput. Vision Graphics and Image Process.* 1993.