PERGAMON

Pattern Recognition 33 (2000) 787–807

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

An adaptive logical method for binarization of degraded document images

Yibing Yang*, Hong Yan

School of Electrical and Information Engineering, University of Sydney, NSW 2006, Australia

Received 29 October 1998; accepted 29 March 1999

Abstract

This paper describes a modified logical thresholding method for binarization of seriously degraded and very poor quality gray-scale document images. This method can deal with complex signal-dependent noise, variable background intensity caused by nonuniform illumination, shadow, smear or smudge and very low contrast. The output image has no obvious loss of useful information. Firstly, we analyse the clustering and connection characteristics of the character stroke from the run-length histogram for selected image regions and various inhomogeneous gray-scale backgrounds. Then, we propose a modified logical thresholding method to extract the binary image adaptively from the degraded gray-scale document image with complex and inhomogeneous background. It can adjust the size of the local area and logical thresholding level adaptively according to the local run-length histogram and the local gray-scale inhomogeneity. Our method can threshold various poor quality gray-scale document images automatically without need of any prior knowledge of the document image and manual fine-tuning of parameters. It keeps useful information more accurately without overconnected and broken strokes of the characters, and thus, has a wider range of applications compared with other methods. © 2000 Pattern Recognition Society. Published by Elsevier Science Ltd. All rights reserved.

Keywords: Document images; Image thresholding; Image segmentation; Image binarization; Adaptive logical thresholding

1. Introduction

Document images, as a substitute of paper documents, mainly consist of common symbols such as handwritten or machine-printed characters, symbols and graphics. In many practical applications, we only need to keep the content of the document, so it is sufficient to represent text and diagrams in binary format which will be more efficient to transmit and process instead of the original gray-scale image. It is essential to threshold the document image reliably in order to extract useful information and make further processing such as character recognition and feature extraction, especially for those poor quality document images with shadows, nonuniform illumination, low contrast, large signal-dependent noise, smear and smudge. Therefore, thresholding a scanned gray-scale image into two levels is the first step and also a critical part in most document image analysis systems since any error in this stage will propagate to all later phases.

Although many thresholding techniques, such as global [1–4] and local thresholding [5–7] algorithms, multi thresholding methods [8–11] and adaptive thresholding techniques [12,13] have been developed in the past, it is still difficult to deal with images with very low quality. Most common problems in poor quality document images are: (1) variable background intensity due to nonuniform illumination and unfit storage, (2) very low local contrast due to smear or smudge and shadows in the capturing process of the document image, (3) poor writing or printing quality, (4) serious signal-dependent noise, and (5) gray-scale changes in highlight and color areas. It is essential to find thresholding methods which can correctly keep all useful information and remove noise and background. Meanwhile, most document

* Corresponding author. Tel.: + 61-2-9351-6210; fax: + 61-2-9351-3847.

E-mail address: ybyang@ee.usyd.edu.au (Y. Yang)

PII: S0031-3203(99)00094-1

processing systems need to process a large number of documents with different styles and layouts every day, thus, they require that whole processing procedure is achieved automatically and adaptively without prior knowledge and pre-specified parameters. Global thresholding methods cannot meet these requirements, and local, or adaptive thresholding methods, which need to be tuned with different parameters according to different image classes, cannot be used for automated process either.

In this paper, we propose a thresholding method based on adaptive logical level technique to binarize seriously degraded and very poor quality gray-scale document images. Our method can deal with complex signal-dependent noise, variable background intensity caused by non-uniform illumination, shadow, smear or smudge and very low contrast without obvious loss of useful information.

The paper is organized as follows. Section 2 briefly reviews related works on image thresholding techniques with an emphasis on the document image binarization based on local analysis and adaptive thresholding. Section 3 analyses various factors which can cause poor quality and inhomogeneous gray-level background in an image and propose a rule to select the local area for analysis and to produce run-length histograms and to extract stroke width information of a document image. Section 4 describes the principle and implementation process of our modified adaptive logical level technique to threshold various degraded and poor quality document images, and a simple and effective method for post processing of binary image. Section 5 discusses and evaluates the experimental results of the proposed method by comparison with some related thresholding techniques according to implementation complexity, character size and stroke width restriction, the number of pre-specified parameters and their meanings and setting, and human subjective evaluation of thresholded images, with experiments on some typical poor quality document images with bad illuminating condition (Fig. 1), and with shadows and signal-dependent noise (Figs. 2 and 3). The last section includes the summary and conclusion of our work.

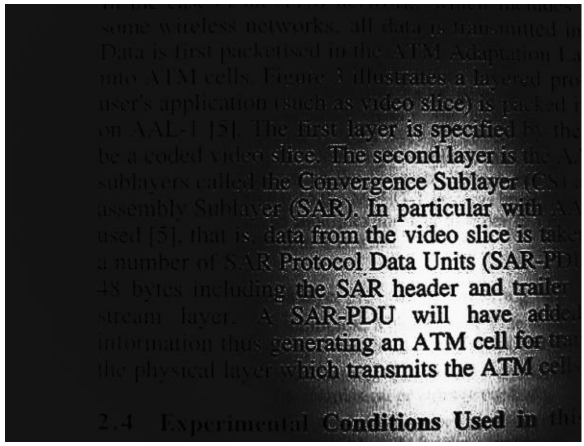


Fig. 1. A $768 \times 576 \times 8$ original document image under bad illuminating condition.

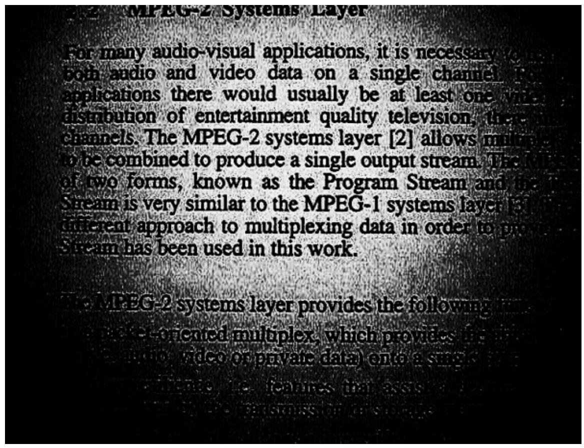


Fig. 2. A $768 \times 576 \times 8$ original document image under bad illuminating condition and signal-dependent noise.

2. Related work

We briefly review some related works on image thresholding, particularly for poor quality document image binarization, which will be evaluated and compared with our thresholding method later. More complete reviews of image thresholding techniques can be found in [2,4,14–16].

Image binarization methods can be divided into two classes: global and local thresholding techniques. The simplest and earliest method is the global thresholding technique. The most commonly used global thresholding techniques are based on histogram analysis [1,3,4]. Threshold is determined from the measure that best separates the levels corresponding to the peaks of the histogram, each of which corresponds to image pixels of a different part like background or objects in the image. Some global multi-threshold techniques are based on edge analysis [9,10] and histogram distribution function [8,11]. Sahoo et al. [2] analysed and evaluated the performance of over 20 popular global thresholding algorithms. All these algorithms need to have a priori knowledge of the image processed about the number of peaks in the gray-level histogram. The modality of the document image histogram, however, may change from image to image. Thus, an obvious drawback of these global techniques is that it cannot separate those areas which

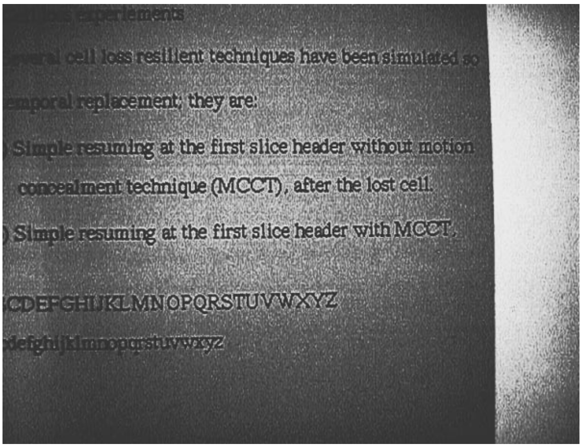


Fig. 3. A $768 \times 576 \times 8$ original document image under bad illuminating condition and noise and shadow.

have the same gray level but do not belong to the same part. These methods do not work well for the document images with shadows, inhomogeneous backgrounds, complex background patterns and different types of fonts and typesettings, which may have a histograms that contains a single peak (Fig. 4). In this case, a single threshold or some multilevel thresholds could not result in an accurate binary document image as shown in Figs. 5–7 no matter how to tune the threshold parameters.

In local and adaptive thresholding techniques, local threshold levels are determined by optimizing some local statistical measures of separation. The criterion function may include the local intensity change (max/min and contrast) [13], stroke width of the characters [17], spacial measures like connectivity and clustering [18,19] and some gradient and edge information [12,20,21]. For complex document image analysis, Kamel and Zhao [22] compared four local adaptive thresholding algorithms for document images with shadows and complex background patterns and proposed two new thresholding techniques – logical level technique and mask-based subtraction technique. Trier and Jain [14,15] evaluated 11 popular local thresholding methods and four global thresholding techniques. For all local thresholding techniques, it appears that none could threshold all images well with a set of operating parameters. In the following, we review a few related local thresholding algorithms particularly for poor quality document image with shadow, signal-dependent noise and inhomogeneous background and their results, which will be compared with our method later.

2.1. Connectivity-based thresholding algorithm

This algorithm was proposed in Ref. [18]. It uses the local connectivity as an information measure. The objective of thresholding is to preserve connectivity within local regions. This algorithm is implemented in three steps.

(1) Determine a histogram of the number of horizontal and vertical runs that result from thresholding the original image at each intensity level. It is equivalent to count all black and white runs along all rows and columns for all binary images corresponding to each intensity level.
(2) Calculate the "sliding profile" from the runs histogram to find plateaus or lack of variation of runs, some ranges around each intensity level can be determined.
(3) Determine the number of thresholds as the number of peaks on the sliding profile. The thresholds are chosen at the peaks that the sliding profile have local maximum values. The image are thresholded into $n + 1$ intensity levels by the n thresholds.

This algorithm produces global thresholds, but uses local connectivity information. It can be used for local thresholding if multi-thresholds are used in different areas of the image. It could not segment the document images well which are badly illuminated, especially when they contain both shadows and noise, as the shadow itself can be regarded as a connected part and the noise can affect the run histogram. We tested this algorithm for some poor quality document images. Some results are shown in Section 5 (from Figs. 19–21).

2.2. Local intensity gradient method (LIG)

This method as presented in Ref. [20] and evaluated and slightly modified in Ref. [21] is based on the principle that objects in an image provide high spatial frequency components and illumination consists mainly of lower spatial frequencies. It first detects edges, and then the interior of objects between edges is filled. First, for each pixel (x, y) in the input image $f(x, y)$, calculate

$$d(x, y) = \min_{i=1,\dots,8} [f(x, y) - f(x_i, y_i)],$$

where (x_i, y_i), $i = 1, \dots, 8$ are the 8-connected neighbours of (x, y). Then the image $d(x, y)$ of minimum local difference is broken up into the regions of size $N \times N$. For each region, the mean m and standard deviations σ are computed. Both values are smoothed by weighted mean and then bilinearly interpolated to produce two new images M and S from m and σ, respectively. Then for all pixels (x, y), if $M(x, y) \geqslant m_0$ or $S(x, y) < \sigma_0$, the pixel is regarded as part of a flat region and remains unlabeled, else, if $d(x, y) < M(x, y) + k(x, y)$, then (x, y) is labeled as print; else (x, y) remains unlabeled. The resulted binary image highlights the edges. This is followed by pixel aggregation and region growing steps to locate the remaining parts of the print objects. This method needs three predetermined

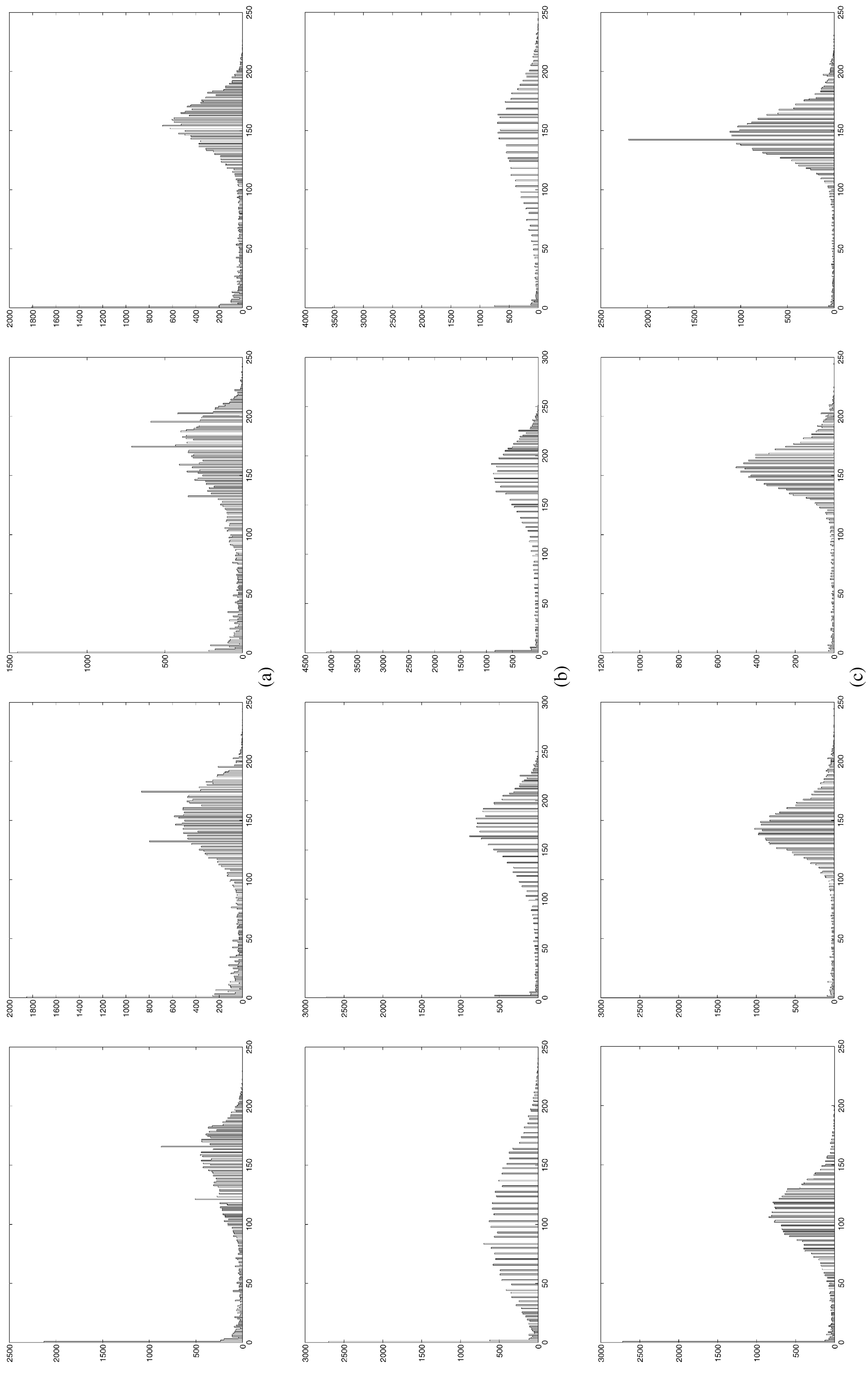


Fig. 4. Some local histograms. (a), (b) and (c) correspond to the local histograms of Figs. 1–3, respectively.

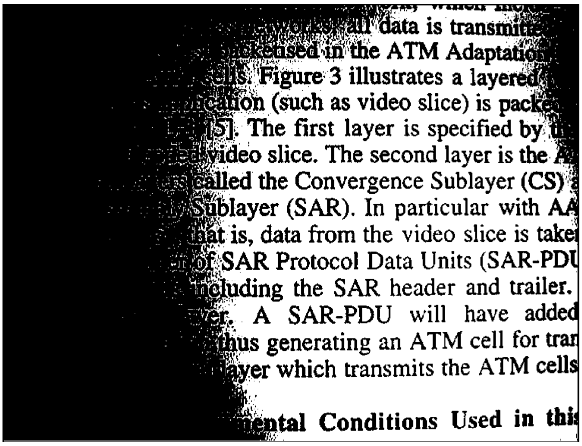

all data is transmitt
ketised in the ATM Adaptatio
cells. Figure 3 illustrates a layered
cation (such as video slice) is pack
[5]. The first layer is specified by
video slice. The second layer is the A
called the Convergence Sublayer (CS)
Sublayer (SAR). In particular with AA
that is, data from the video slice is take
of SAR Protocol Data Units (SAR-PD
including the SAR header and trailer.
er. A SAR-PDU will have added
thus generating an ATM cell for tra
layer which transmits the ATM cells
ental Conditions Used in thi

Fig. 5. Binary document image extracted using the global method from the original image of Fig. 1.

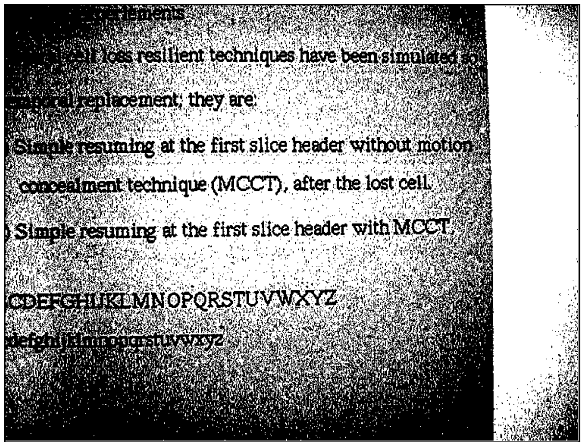

cell loss resilient techniques have been simulated
replacement; they are:
Simple resuming at the first slice header without motion
concealment technique (MCCT), after the lost cell.
Simple resuming at the first slice header with MCCT.
CDEFGHIJKLMNOPQRSTUVWXYZ

Fig. 7. Binary document image extracted using the global method from the original image of Fig. 3.

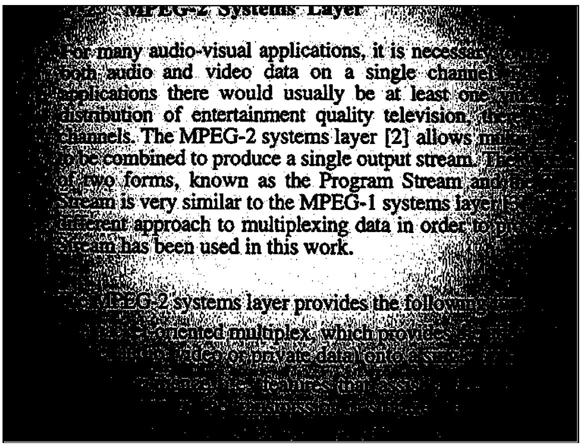

MPEG-2 Systems Layer
For many audio-visual applications, it is necessary
both audio and video data on a single channel
applications there would usually be at least one
distribution of entertainment quality television,
channels. The MPEG-2 systems layer [2] allows
to be combined to produce a single output stream.
of two forms, known as the Program Stream and
Stream is very similar to the MPEG-1 systems layer
different approach to multiplexing data in order to
Stream has been used in this work.
MPEG-2 systems layer provides the following

Fig. 6. Binary document image extracted using the global method from the original image of Fig. 2.

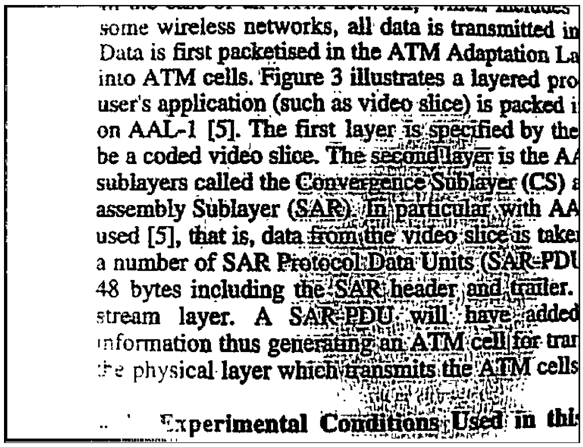

some wireless networks, all data is transmitted in
Data is first packetised in the ATM Adaptation La
into ATM cells. Figure 3 illustrates a layered pro
user's application (such as video slice) is packed i
on AAL-1 [5]. The first layer is specified by the
be a coded video slice. The second layer is the A
sublayers called the Convergence Sublayer (CS) a
assembly Sublayer (SAR). In particular with AA
used [5], that is, data from the video slice is take
a number of SAR Protocol Data Units (SAR-PDU
48 bytes including the SAR header and trailer.
stream layer. A SAR-PDU will have added
information thus generating an ATM cell for tran
the physical layer which transmits the ATM cells
Experimental Conditions Used in thi

Fig. 8. Binary document image extracted using local intensity gradient method with $N = 16$, $m_0 = -1.0, \sigma_0 = 1.0$ and $k = -1.0$ from the original image in Fig. 1.

parameters m_0, σ_0 and k and block size N. We tested this method for several images with $N = 16, m_0 = -1.0$, $\sigma_0 = 1.0$ and $k = -1.0$. The results are shown in Figs. 8–10. It can deal with slowly changing background with bad illumination. It will, however, intensify some noise effects and could not work well for fast changing background with bad illumination due to gradient-based analysis.

2.3. *Intergrated function algorithm and its improvement*

This technique as described in Ref. [12] and as improved and evaluated in Refs. [14,21] applies a gradient-like operator, defined as the activity $A(x, y)$, which is the absolute sum of approximated derivatives for both scan and raster directions taken over a small area, on the image. Pixels with activity below a predetermined threshold T_a are labelled '0'. The other pixels are further tested by the *Laplacian* edge operator $dd_{xy}(x, y)$. The pixel is labelled '+' if $dd_{xy}(x, y) > 0$; otherwise '−'. Thus, a three-level label-image with pixel levels '+', '0' and '−' is produced. In a sequence of labels along with some straight line passing through the currently processed points (x, y), edges are identified as '− +' or '+ −' transitions. Object pixels are assumed to be '+' and '0' labelled pixels between a '− +' and '+ −' pair. The distance between this pair can be regarded as the "strokewidth" along this line for document images. Background pixels tend not to be included between this pair.

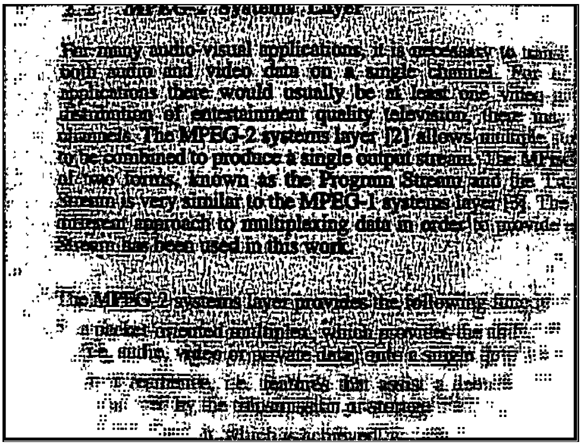


Fig. 9. Binary document image extracted using local intensity gradient method with $N = 16, m_0 = -1.0, \sigma_0 = 1.0$ and $k = -1.0$ from the original image in Fig. 2.

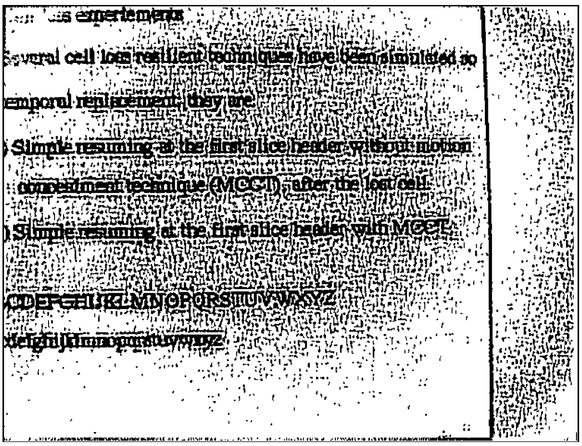


Fig. 10. Binary document image extracted using local intensity gradient method with $N = 16, m_0 = -1.0, \sigma_0 = 1.0$ and $k = -1.0$ from the original image in Fig. 3.

According to this analysis, a 2×2 region is classified at a time, it is needed that all the four pixels are inside either horizontal or vertical object pixel sequences. Trier [21] improved this algorithm mainly in that all '+' marked regions are labelled print, and '−' marked regions are labeled background; a '0' marked region is labeled print if a majority of the pixels with 4-connected are '+' marked, otherwise it is labelled background. It is sensitive to noises and fast changing background due to the *Laplacian* edge operator. Some thresholding results using this algorithm are shown in later section on the experiment results of this paper (from Figs. 22–24).

2.4. Local contrast technique

Giuliano et al. [23] presented the local contrast technique to extract binary image in their patent for a character recognition system. This technique is implemented in a 9×9 window for an input image $f(x, y)$. Each pixel in the output binary image $b(x, y)$ is determined on the $3\times3\times5$ local pixels within a 9×9 window as shown in Fig. 11. We use gray level 1 to represent foreground (print) and 0 to represent background (no print) in an output binary image. This method can be implemented as follows:

if $f(x, y) < T_1$, then, $b(x, y) = 1$;
otherwise, $A_{2t} = \{(x, y) | (x, y) \in A_2 \text{ and } f(x, y) > T_2\}$;
$a_1 =$ mean of 9 pixels in area A_1; $a_2 =$ mean of the pixels in area A_{2t};
if $T_3 a_2 + T_5 > T_4 a_1$, then, $b(x, y) = 1$;
otherwise, $b(x, y) = 0$.

where, T_1–T_5 are five predetermined parameters, T_1 is equivalent to the threshold in the global technique, T_2 is used to detect all pixels in A_2 with gray levels over T_2 itself, other parameters are used to compare the mean a_1 of the central region A_1 of processed pixel with the mean a_2 of the pixels over T_2 in the four corner regions. The biggest difficulty of this method is how to choose predetermined parameters. Different parameter setting could produce quite different results. This method is sensitive to inhomogeneous background, large shadows and noise. Figs. 25–27 in Section 5 show some tested results produced by this method.

The above four methods will be compared with our thresholding method.

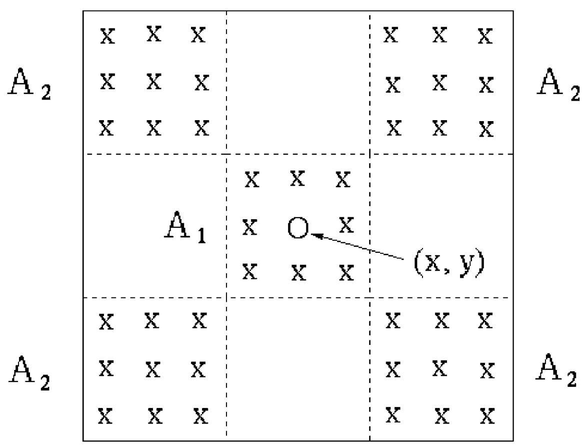


Fig. 11. Neighbour analysis in the local contrast technique.

3. Document image background and stroke width analysis

For some poor quality document images with variable or inhomogeneous background intensity like shadows, smear or smudge, complex background patterns and signal-dependent noise, a practical problem is that no thresholding algorithm could work well for all kinds of document images. Most commonly, some methods or some parameters for the document image with variable or inhomogeneous background intensity and noises will result in a thresholded image in which printed characters have nonuniform stroke width and possibly even lost strokes or false character and connection caused by background noise as shown in Figs. 8 and 9, 28(a) and 29(a) in Section 5. This will result in low character recognition rate and document image compression rate in later processing. Background and stroke width analysis of the characters for the document image can overcome or reduce this problem and improve thresholding accuracy and robustness. Here, we present a simple and efficient method for the backgroundand character stroke width analysis.

3.1. Background analysis

When an image consists of only objects and the background, the best way to pick up a threshold is to search a histogram, assuming it is bimodal, and find a gray level which separates the two peaks. However, problems arise when the object area is small compared to the background area or when both the object and the background assume some broad range of gray levels as the background gray-level distributions and character gray-level distributions. Two examples are shown in Figs. 12 and 13. In these cases, the histogram is no longer bimodal as shown in Fig. 4. But, in some local areas, the bimodality of the local histogram could be more obvious if this area

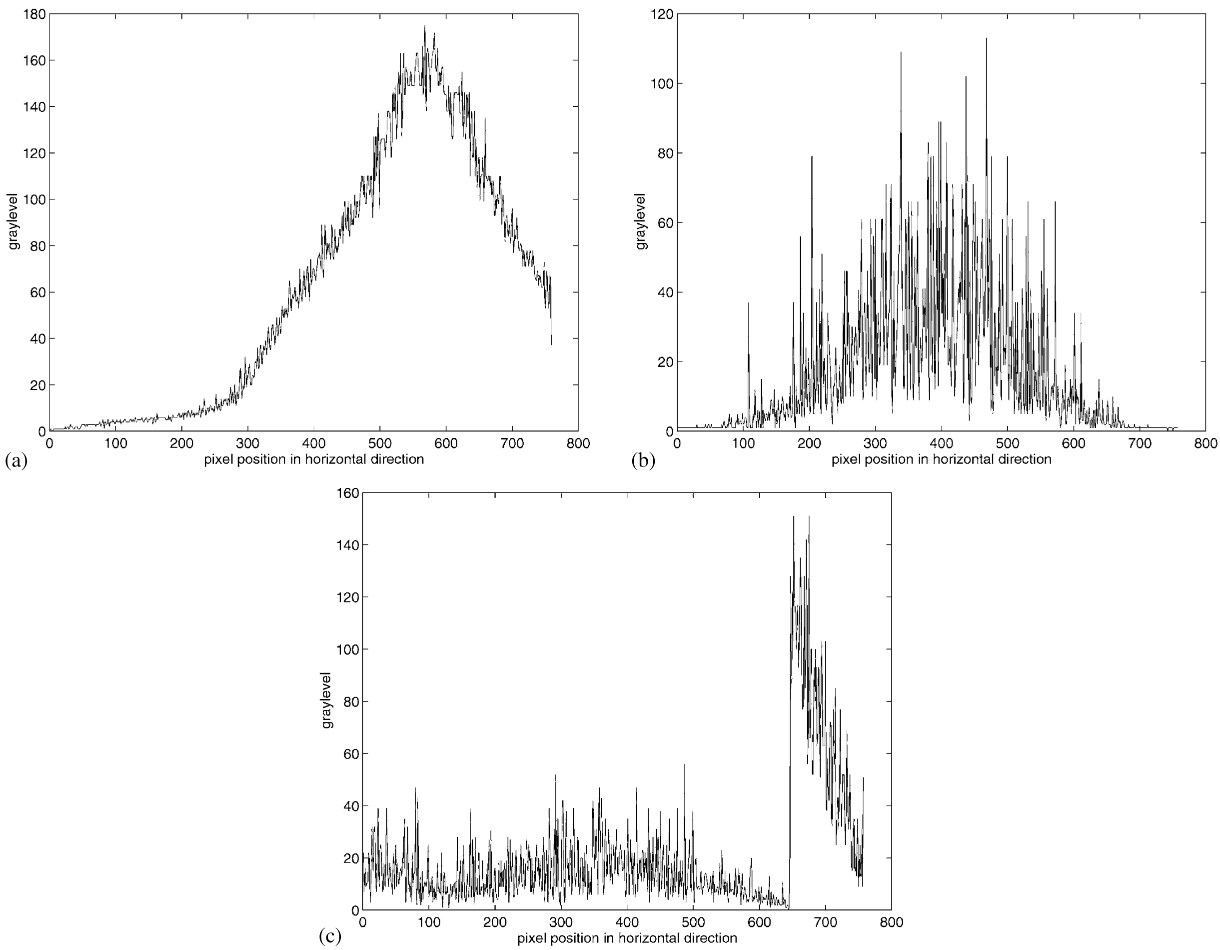


Fig. 12. Examples of gray-scale distribution of document image backgrounds with bad illuminated condition and signal-dependent noise. (a), (b) and (c) correspond to the backgrounds of the document images in Figs. 1–3, respectively.

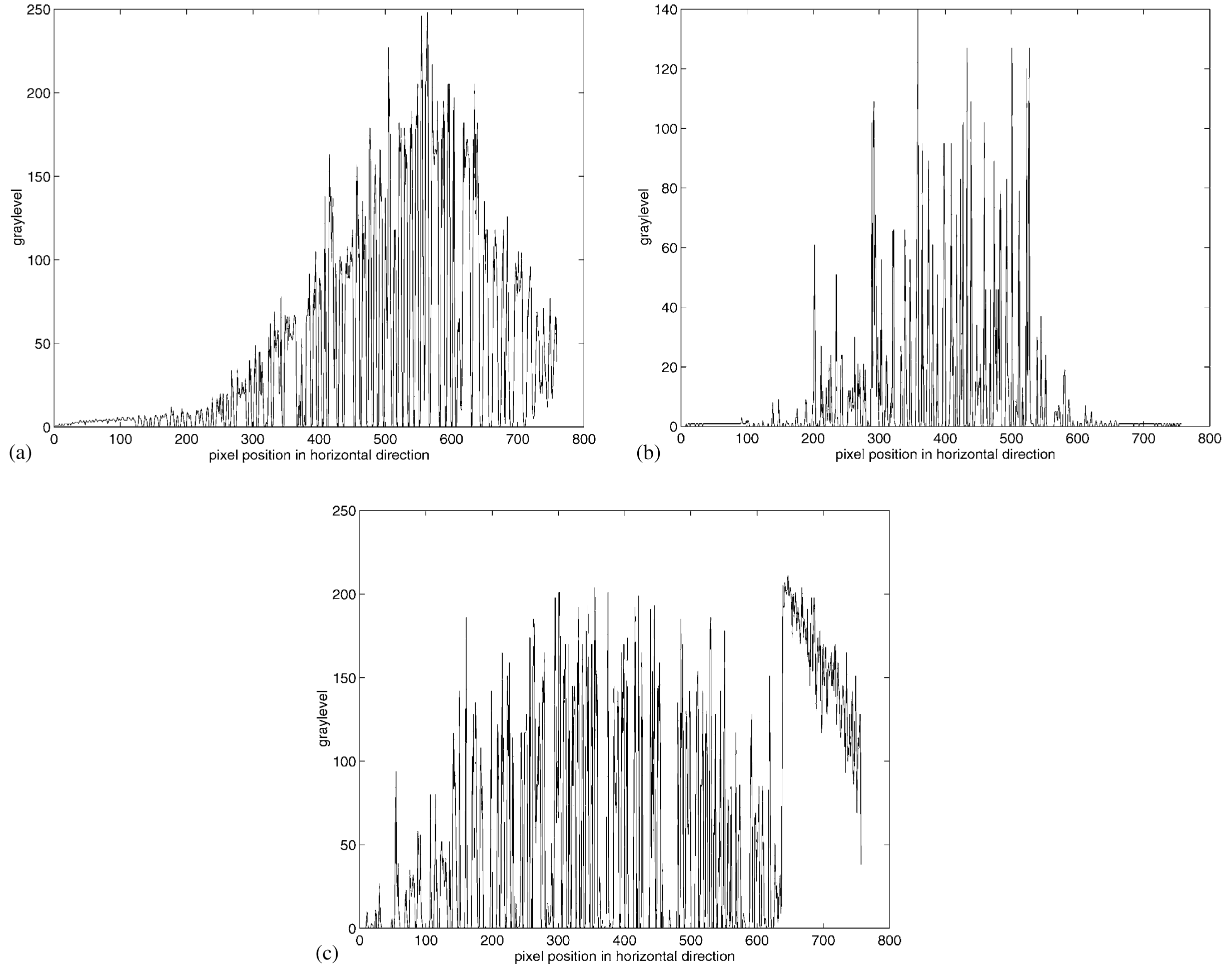


Fig. 13. Examples of gray-scale distribution of document image foregrounds with bad illuminated condition and signal-dependent noises. (a), (b) and (c) correspond to the gray-scale distribution for the character line of the document images in Figs. 1–3, respectively.

contains separatable background and objects/characters. We divide an image into $N \times N$, $(N = 4, \ldots, 8)$ regions in order to find some local areawith quasi-bimodal local histogram or higher local contrast, and then make local histogram analysis for the regions in the two diagonal directions in Fig. 14(a) if N is even and for the regions in the two diagonal, horizontal and vertical directions in Fig. 14(b) if N is odd. We can gradually increase the local regions or the directions of analysis until $N = 8$ if no quasi-bimodal local histogram is found in the case for $N < 8$. In smaller regions, those regions with the same patterns in Fig. 14 are simultaneously analysed in each analysis. From the local region analysis, some local region histograms with quasi-bimodal property are shown in Fig. 15. With local quasi-bimodal histogram, the character stroke widths and background changes can be analysed using run-length histograms from these areas.

3.2. Stroke width and noise analysis

The document image can be accurately thresholded if the average or maximum stroke width of the characters in the document image may be determined, because highly structured-stroke units frequently appear in most document images. We have found some regions with quasi-bimodal local histograms in the poor quality document images by using local region analysis, then, local run-length information can be extracted to form a run-length histogram from those selected local regions with quasi-bimodal local histograms. The stroke width information and background noise can be achieved by analysing run-length histogram. Here, we only analyse those selected image regions and only consider black runs related to the characters or other objects. We denote a run-length histogram as a one-dimensional array

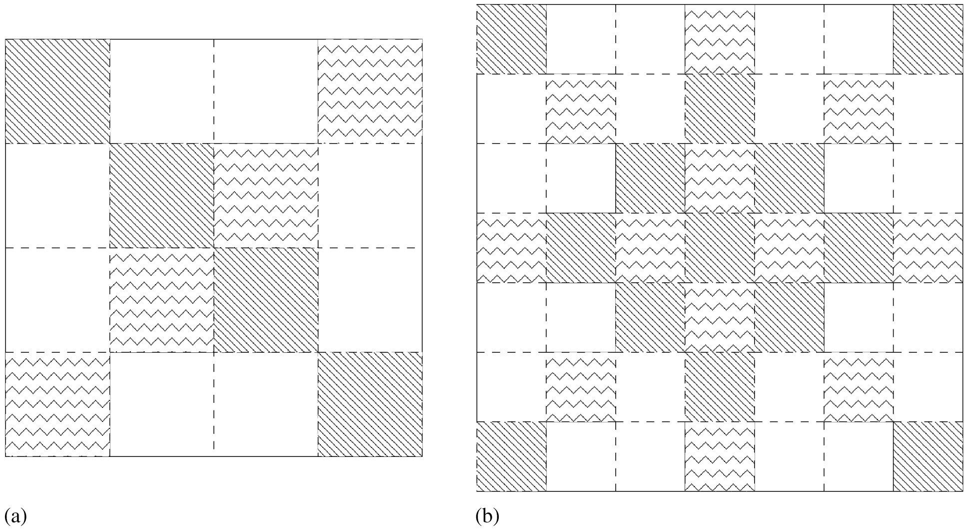


Fig. 14. Local region histogram analysis to find some regions with quasi-bimodal local histogram or higher local contrast. (a) local region analysis in the two diagonal directions; (b) fine local region analysis in the two diagonal, horizontal and vertical directions.

$R(i), i \in I, I = \{1, 2, \ldots, L\}$, where L is the longest run to be counted. $R(i)$ is frequency of the run of length i. Black run-length can be counted from the one-dimensional gray level distributions across the selected local regions as shown in Fig. 16 in horizontal and vertical directions. The number of the directions, which are across the character/object region in the selected image regions with quasi-bimodal local histograms, can be increased to four directions, including horizontal, vertical and two diagonal directions if the document image contains complex symbol patterns.

The *stroke width* (SW) is defined as the run-length with the highest frequency in the run-length histogram excluding the unit run-length. That is, $\mathrm{SW} = i$, if $R_{\max}(i) = \max_{i \in I} R(i), i \neq 1$. It actually reflects the average width of strokes in a document image. Fig. 17 illustrates the run-length histograms, from which, stroke width can be easily determined. If an image contains some complex background patterns or noises, the highest peak may be formed by these factors instead of the characters. In this case, selected region analysis and only black run-length analysis become necessary to prevent producing wrong stroke width. Statistical study shows that the mean stroke width is usually over one pixel, accordingly, all unit-runs should be removed as background in resulting binary image no matter how it is produced by noise or other background changes. We can use the unit-run noise (URN) [17] to measure background noise and changes.

$$\mathrm{URN} = \frac{R(1)}{\max_{i \in I} R(i)}, \quad i \neq 1.$$

A high number of unit runs means that a document image contains high noise background and/or fast changing background.

4. Adaptive logical level thresholding technique

4.1. Logical level technique

Logical level technique proposed by Kamel and Zhao [22] is developed on the basis of analysing integrated function algorithm [12]. It is based on the idea of comparing the gray level of the processed pixel or its smoothed gray level with some local averages in the neighborhoods about a few other neighbouring pixels. More than once the comparison results are regarded as *derivatives*. Therefore, pixel labeling, detection and extraction using the *derivatives*, the logical bound on the ordered sequences and the stroke width range can be adopted. This technique processes each pixel by simultaneously comparing its gray level or its smoothed gray level with four local averages in the $(2\mathrm{SW}+1) \times (2\mathrm{SW}+1)$ window centered at the four points P_i, P'_i, P_{i+1}, P'_{i+1} shown in Fig. 18. We use 1 to represent character/object and 0 to represent background in the resulting binary image. Mathematically, this technique can be described as follows:

$$b(x, y) = \begin{cases} 1 & \text{if } \bigvee_{i=0}^{3}[L(P_i) \wedge L(P'_i) \wedge L(P_{i+1}) \wedge L(P'_{i+1})] \text{ is true,} \\ 0 & \text{otherwise,} \end{cases}$$

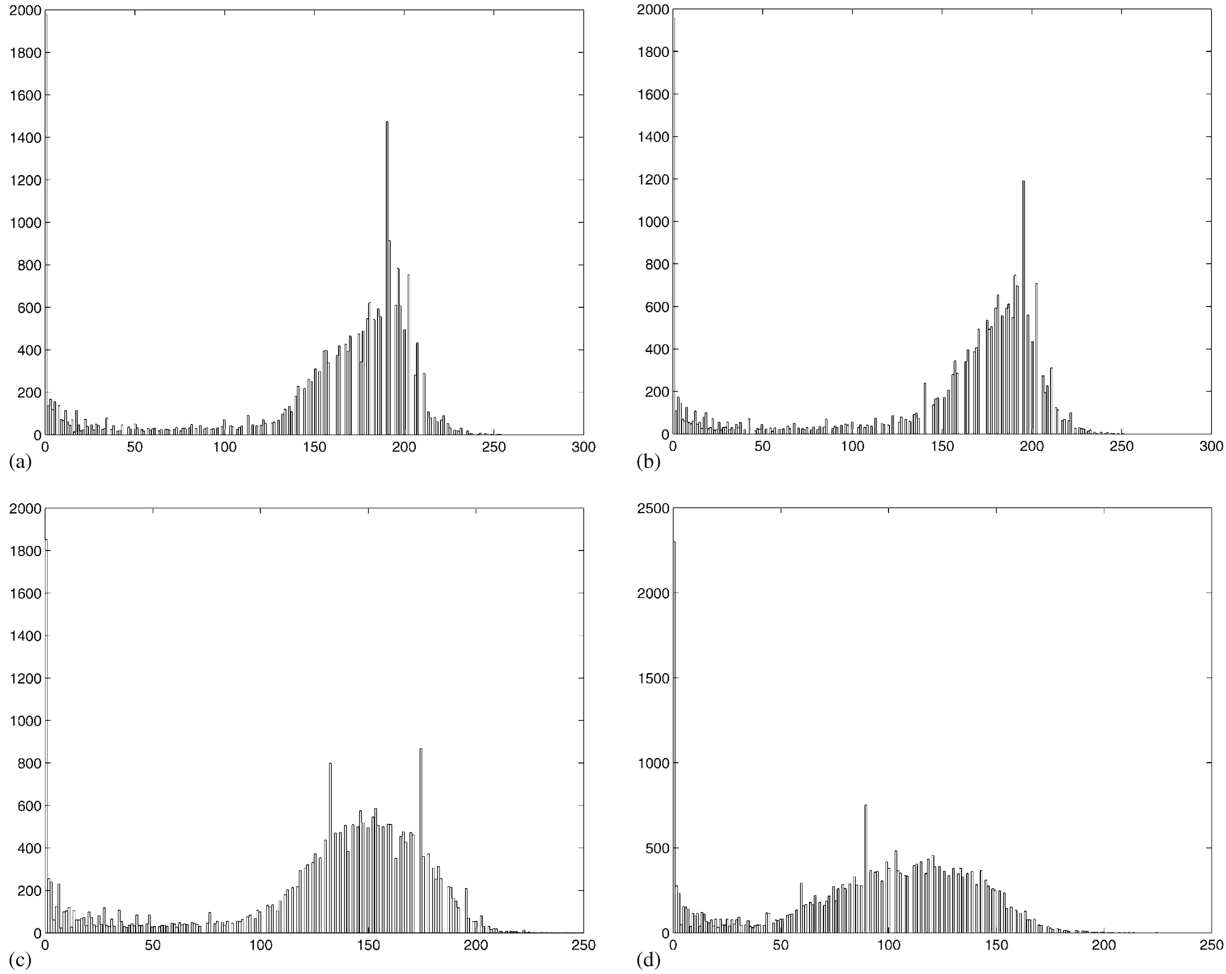


Fig. 15. Some local region histograms with quasi-bimodal property or the regions with larger contrast in the document images with bad illuminated condition and signal-dependent noise. (a) and (b) correspond to some local region histograms of Fig. 1, (c) and (d) correspond to some local region histograms of Fig. 2.

where SW is the predetermined maximal stroke width, $P'_i = P_{(i+4)\mathrm{mod}\ 8}$, for $i = 0,\ldots,7$, $L(P) = ave(P) - g(x, y) > T$, T is a predetermined parameter,

$$ave(P) = \sum_{-\mathrm{SW} \leqslant i \leqslant \mathrm{SW}} \sum_{-\mathrm{SW} \leqslant j \leqslant \mathrm{SW}} f(P_x - i, P_y - j) / (2 \times \mathrm{SW} + 1)^2,$$

P_x, P_y are the coordinates of P and $g(x, y) = f(x, y)$ or its smoothed value.

In order to reduce the computation, fast algorithms are used to calculate the local averages and logical levels.

4.2. Adaptive improvement of logical level technique

We propose some improvements for the original logical level technique in order to achieve automatic and adaptive thresholding and accurate binary images for various poor quality document images. Our modification is made in two aspects. The first one is to determine average maximal stroke width SW automatically by run-length histograms in the selected local regions of the image as described in the preceding section. This stroke width can be tuned automatically for different document images. As usual, we use the run-length with highest peak of the run-length histogram in the selected regions of the document image, $\mathrm{SW} = i$, if $R_{\max}(i) = \max_{i \in I} R(i), i \neq 1$, as stroke width. In some cases, we may also use the run-length $\mathrm{SW}_{2nd} = j$ of the second highest peak $R_{2nd-right-peak}(j)$ right to the highest peak, which is the second high peak on the right of the highest peak in the run-length histogram as the stroke width, if $1 \leqslant (j - i) \leqslant 2, i, j \neq 1$ and $R_{2nd-right-peak}(j)/R_{\max}(i) \geqslant 0.8$.

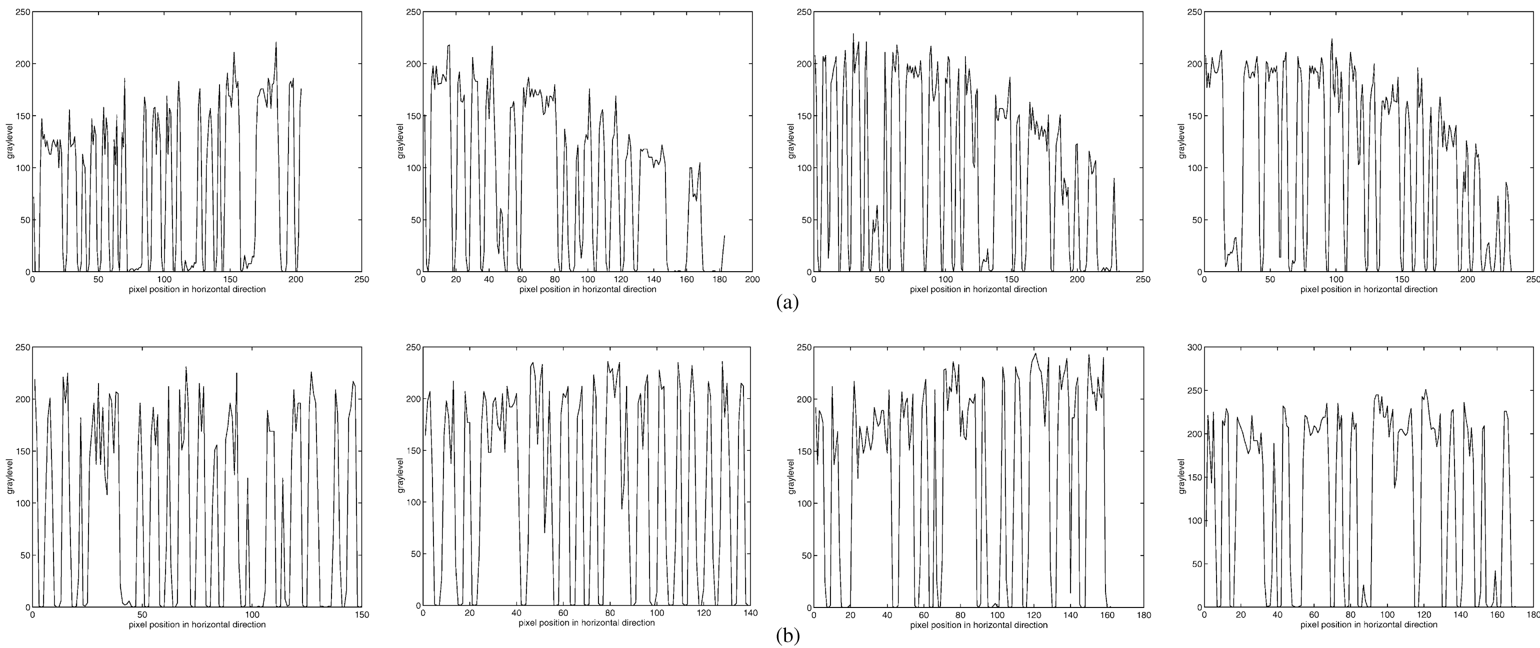
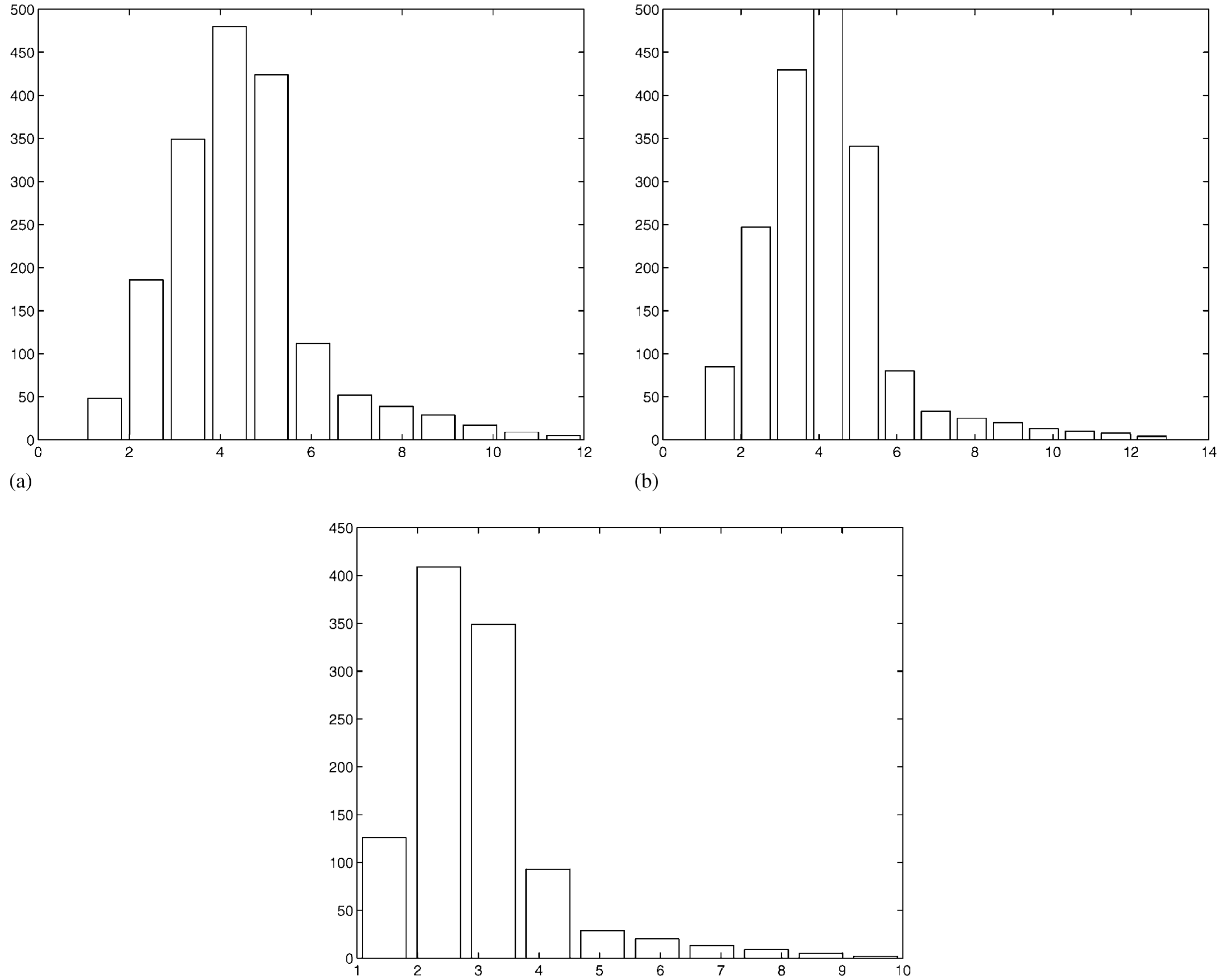


Fig. 16. Some gray-scale distributions across the characters in the selected regions of the document images. We can get the average stroke width information from the run-length of gray-level change. (a) corresponds to gray-scale distributions across the selected region of Fig. 1, (b) corresponds to gray-scale distributions across the selected region of Fig. 2.



(a) (b) (c)

Fig. 17. Local run-length histograms, (a), (b) and (c) correspond to the run-length histograms of the selected regions from the document images in Figs. 1–3, respectively.

The other improvement is to automatically and adaptively produce local parameter T instead of using predetermined global parameter. It can overcome the uneven thresholding effect and false thresholding result for the document images with bad illumination, inhomogeneous and fast changing background and big noise by the original logical level technique. Parameter T is adaptively and automatically produced as follows:

1. Calculate $f_{\mathrm{SW-max}}(x, y) = \max_{x_i,\ y_i \in (2\mathrm{SW}+1)\ window} f(x_i, y_i)$ and $f_{\mathrm{SW-min}}(x, y) = \min_{x_i,\ y_i \in (2\mathrm{SW}+1)\ window} f(x_i, y_i)$ in the $(2\mathrm{SW}+1) \times (2\mathrm{SW}+1)$ window centered at the processed point P.
2. Calculate $|f_{\mathrm{SW-max}}(x, y) - ave(P)|$ and $|f_{\mathrm{SW-min}}(x, y) - ave(P)|$;
3. If $|f_{\mathrm{SW-max}}(x, y) - ave(P)| > |f_{\mathrm{SW-min}}(x, y) - ave(P)|$, the local $(2\mathrm{SW}+1) \times (2\mathrm{SW}+1)$ window region tends to contain more local low gray levels, then, $T = \alpha(\frac{2}{3} f_{\mathrm{SW-min}}(x, y) + \frac{1}{3} ave(P))$. Here, α can be a fixed value between 0.3 and 0.8. It can be taken as $\frac{1}{3}$ for very poor quality images with high noise and low contrast like in our examples. In most cases, it can be taken as $\frac{2}{3}$.
4. If $|f_{\mathrm{SW-max}}(x, y) - ave(P)| < |f_{\mathrm{SW-min}}(x, y) - ave(P)|$, the local $(2\mathrm{SW}+1) \times (2\mathrm{SW}+1)$ window region tends to contain more local high graylevels, then, $T = \alpha(\frac{1}{3} f_{\mathrm{SW-min}}(x, y) + \frac{2}{3} ave(P))$.
5. If $|f_{\mathrm{SW-max}}(x, y) - ave(P)| = |f_{\mathrm{SW-min}}(x, y) - ave(P)|$;

   ○ If $f_{\mathrm{SW-max}}(x, y) = f_{\mathrm{SW-min}}(x, y)$, expand the window size to $(2\mathrm{SW}+3) \times (2\mathrm{SW}+3)$, then, repeat from

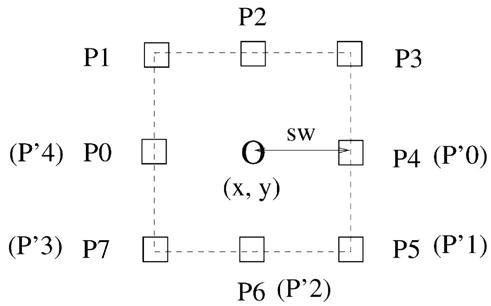


Fig. 18. Processing neighbourhood of logical level thresholding technique.

step 1 but using new widow size. If still $f_{SW-max}(x, y) = f_{SW-min}(x, y)$ in the new window, then, P is regarded as background pixel (or $T = \alpha \cdot ave(P)$).

- If $f_{SW-max}(x, y) \neq f_{SW-min}(x, y)$, the local $(2SW + 1) \times (2SW + 1)$ window region tends to contain the same quota of low and high gray levels, expand the window size to $(2SW + 3) \times (2SW + 3)$, then, repeat from step 1 but using new window size. If $|f_{SW-max}(x, y) - ave(P)| = |f_{SW-min}(x, y) - ave(P)|$ and $f_{SW-max}(x, y) \neq f_{SW-min}(x, y)$ in the new window, then, $T = \alpha \cdot ave(P)$.

4.3. Postprocessing of binary image

The aim of postprocessing for binary image is to remove binary noise and false print information to improve binary quality. In our method, we use run-length information to decide false information. First, the run-length histograms only for print information of the binary image in horizontal and vertical directions are extracted, we compare them with the local run-length histograms from the original document image, the unit-run parts in both horizontal and vertical directions are removed. Then, those runs only with unit and two pixel width combined in both horizontal and vertical directions are removed. Furthermore, we analyse possibly big false print information caused by fast changing backgrounds, which we call long-run noise. A run is considered to be long if it is substantially longer than the maximum run-length of the characters. The number of long runs should be quite small even if underlines, tables and graphics exist in document images. Here, we use a long-run noise (*LRN*) feature [17] to describe if there are some long-run noises in the resulting binary image.

$$\mathrm{LRN} = \frac{\sum_{i > L_0} R(i)}{\max_{i \in I} R(i)}, \quad i \neq 1,$$

where L_0 is a constant, it may be determined larger than average character size. We use LRN to measure if the binary image contains more long-run noise. If it is larger than 1 or close to 1, we will revise the width parameter to smaller in thresholding so that the resulting binary image may be more clear. This process is automatic.

5. Experimental results and evaluation

We have tested six local adaptive thresholding algorithms including the logical level technique and our modified logical technique for a number of document images with poor quality such as with bad illumination, shadow, signal-dependent noise and various variable backgrounds in different parameters. All algorithms were implemented and tested through software written in C programming language in the UNIX on a Sun Sparc Station IPX. Figs. 8–10 and 19–38 illuminate the experiment results respectively using Local Intensity Gradient Method, Connectivity-Based Thresholding Algorithm, Intergrated Function Algorithm, Local Contrast Technique, Logical Level Technique and our Adaptive Logical Thresholding Algorithm. Table 1 gives the average computation time (CPU time in seconds) for the algorithms mentioned above. All images tested are with width of 768, height of 576 and gray-level range of [0,255].

Connectivity-Based Thresholding Algorithm could not segment badly illuminated document images well, especially when they contain both shadows and noises as in Figs. 19–21, since the shadow itself can be regarded as a connected part and the noise could give a big effect for the run histogram. Moreover, its efficiency of implementation is limited by a large number of calculations and decisions, since it needs to calculate the run-length

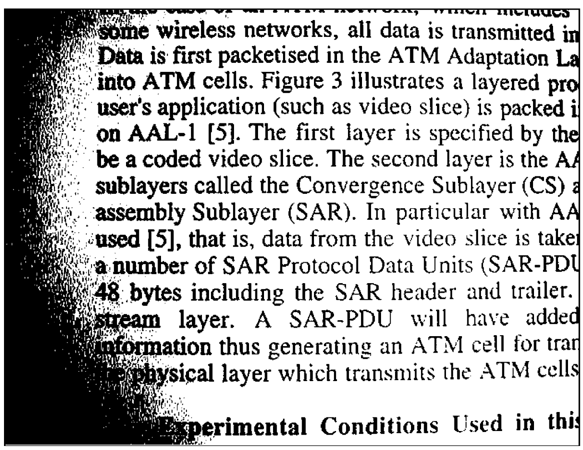
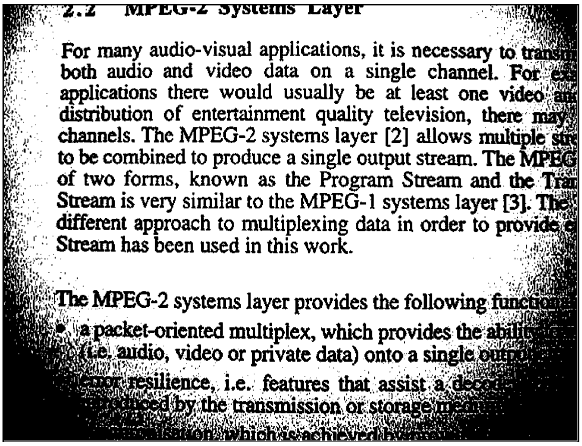

some wireless networks, all data is transmitted in
Data is first packetised in the ATM Adaptation La
into ATM cells. Figure 3 illustrates a layered pro
user's application (such as video slice) is packed i
on AAL-1 [5]. The first layer is specified by the
be a coded video slice. The second layer is the A
sublayers called the Convergence Sublayer (CS) a
assembly Sublayer (SAR). In particular with AA
used [5], that is, data from the video slice is take
a number of SAR Protocol Data Units (SAR-PDU
48 bytes including the SAR header and trailer.
stream layer. A SAR-PDU will have added
information thus generating an ATM cell for tran
physical layer which transmits the ATM cells
Experimental Conditions Used in thi

Fig. 19. Binary document image extracted using connectivity-based thresholding algorithm from the original document image in Fig. 1.


For many audio-visual applications, it is necessary to
both audio and video data on a single channel. For
applications there would usually be at least one video
distribution of entertainment quality television, there
channels. The MPEG-2 systems layer [2] allows multiple
to be combined to produce a single output stream. The
of two forms, known as the Program Stream and the
Stream is very similar to the MPEG-1 systems layer [3]. The
different approach to multiplexing data in order to provide
Stream has been used in this work.

The MPEG-2 systems layer provides the following
• a packet-oriented multiplex, which provides the
(i.e. audio, video or private data) onto a single

Fig. 20. Binary document image extracted using connectivity-based thresholding algorithm from the original document image in Fig. 2.

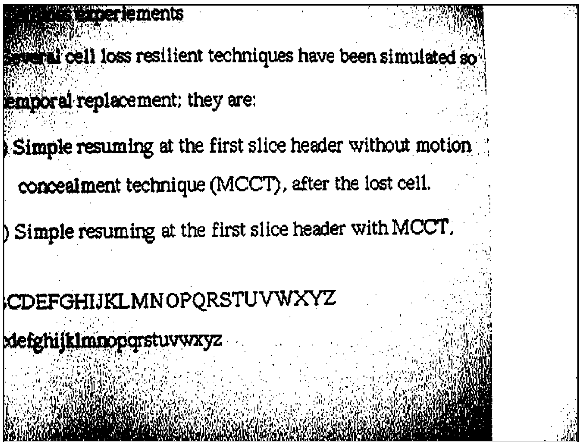

experiements
cell loss resilient techniques have been simulated so
emporal replacement; they are:
Simple resuming at the first slice header without motion
concealment technique (MCCT), after the lost cell.
Simple resuming at the first slice header with MCCT,
CDEFGHIJKLMNOPQRSTUVWXYZ
defghijklmnopqrstuvwxyz

Fig. 21. Binary document image extracted using connectivity-based thresholding algorithm from the original document image in Fig. 3.

histograms for two directions (horizontal and vertical) at each intensity level. Local Intensity Gradient Method can work well with slowly changing background and bad illumination. It will, however, intensify some noise effects and could not work well for fast changing background with bad illumination as shown in Figs. 8–10 due to gradient-based analysis. Besides, the calculation of local minimum difference image (for each pixel) and local mean m and standard deviation σ particularly with the increase of the blocked region size N are quite time-consuming. The selection of pre-specified parameters are image-directed and region-directed. Different pre-specified parameters and region size could result in quite different result images.

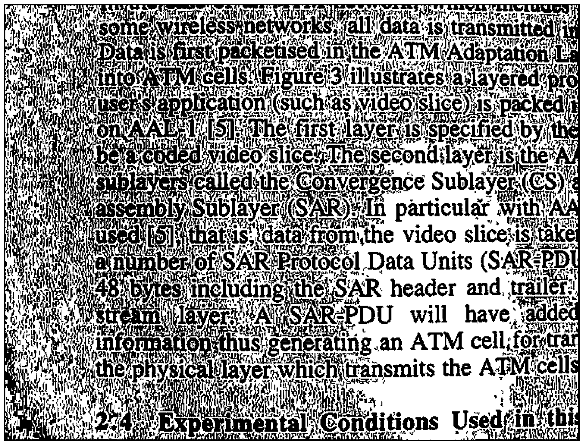

some wireless networks, all data is transmitted
Data is first packetised in the ATM Adaptation
into ATM cells. Figure 3 illustrates a layered
user's application (such as video slice) is packed
on AAL-1 [5]. The first layer is specified by the
be a coded video slice. The second layer is the
sublayers called the Convergence Sublayer (CS)
assembly Sublayer (SAR). In particular with
used [5], that is, data from the video slice is
a number of SAR Protocol Data Units (SAR-PDU
48 bytes including the SAR header and trailer.
stream layer. A SAR-PDU will have added
information thus generating an ATM cell for
the physical layer which transmits the ATM cells
2.4 Experimental Conditions Used in this

Fig. 22. Binary document image extracted using intergrated function algorithm from the image in Fig. 1.

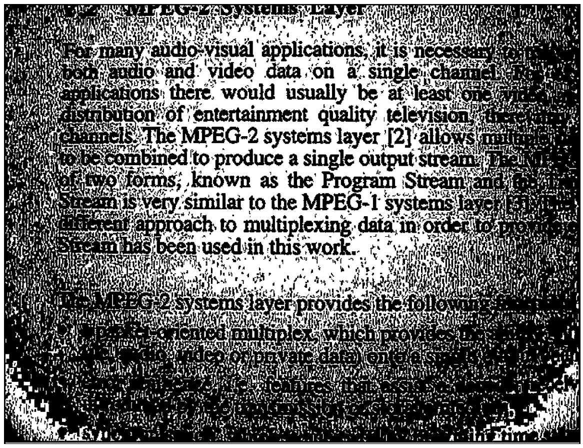

For many audio-visual applications, it is necessary to
both audio and video data on a single channel.
applications there would usually be at least one
distribution of entertainment quality television,
channels. The MPEG-2 systems layer [2] allows
to be combined to produce a single output stream.
of two forms, known as the Program Stream and
Stream is very similar to the MPEG-1 systems layer
different approach to multiplexing data in order to
Stream has been used in this work.

Fig. 23. Binary document image extracted using intergrated function algorithm from the image in Fig. 2.

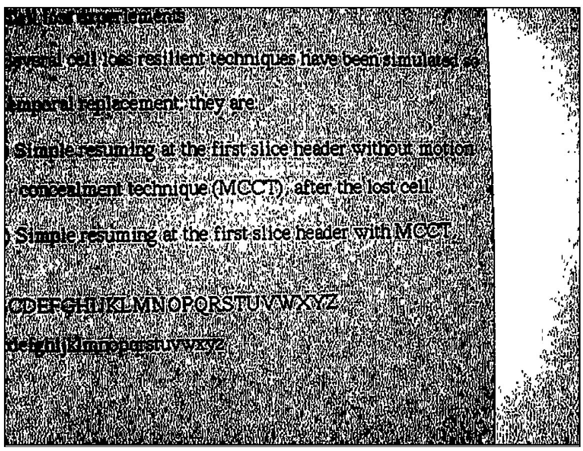

cell loss resilient techniques have been simulated so
emporal replacement; they are:
Simple resuming at the first slice header without motion
concealment technique (MCCT), after the lost cell.
Simple resuming at the first slice header with MCCT,
CDEFGHIJKLMNOPQRSTUVWXYZ

Fig. 24. Binary document image extracted using intergrated function algorithm from the image in Fig. 3.

Intergrated Function Algorithm has fully considered the stroke width information so that it can remove all the large dark areas completely. The labeling and logical detecting assure that every large dark area is a connected black blob which is removed from the image in the final extraction phase. Therefore, the resulting images have no unwanted edges of large dark areas, but it is sensitive to noises and fast changing background due to using *Laplacian* edge operator. It could produce some small noise

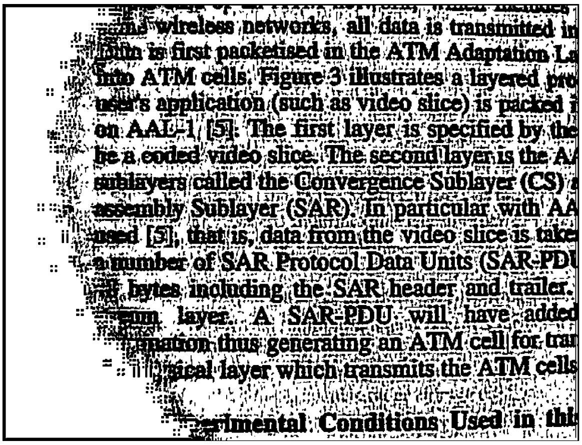

wireless networks, all data is transmitted in
is first packetised in the ATM Adaptation La
into ATM cells. Figure 3 illustrates a layered pro
user's application (such as video slice) is packed i
on AAL-1 [5]. The first layer is specified by the
be a coded video slice. The second layer is the AA
sublayers called the Convergence Sublayer (CS) a
assembly Sublayer (SAR). In particular with AA
used [5], that is, data from the video slice is take
a number of SAR Protocol Data Units (SAR-PDU
bytes including the SAR header and trailer.
layer. A SAR-PDU will have added
mation thus generating an ATM cell for tra
sical layer which transmits the ATM cells
rimental Conditions Used in thi

Fig. 25. Binary document image extracted using local contrast analysis from the image in Fig. 1.

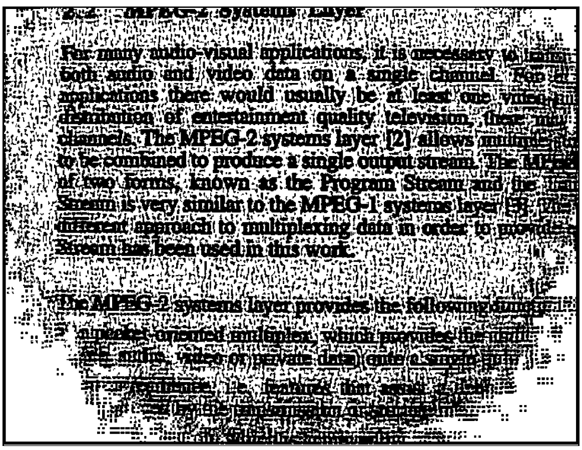

For many audio-visual applications, it is necessary to
both audio and video data on a single channel.
applications there would usually be at least one video
distribution of entertainment quality television,
channels. The MPEG-2 systems layer [2] allows
to be combined to produce a single output stream. The MPEG
of two forms, known as the Program Stream and the
Stream is very similar to the MPEG-1 systems layer
different approach to multiplexing data in order to
Stream has been used in this work.
The MPEG-2 systems layer provides the following

Fig. 26. Binary document image extracted using local contrast analysis from the image in Fig. 2.

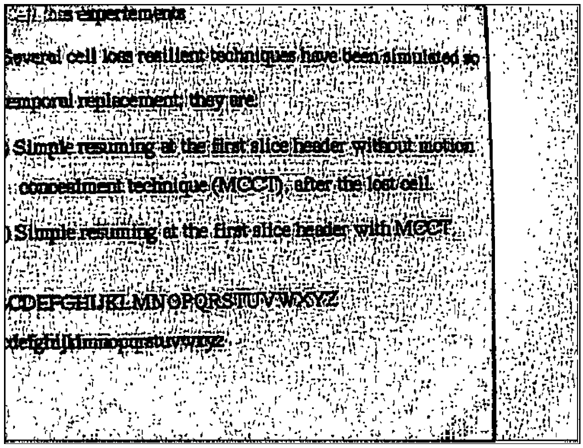
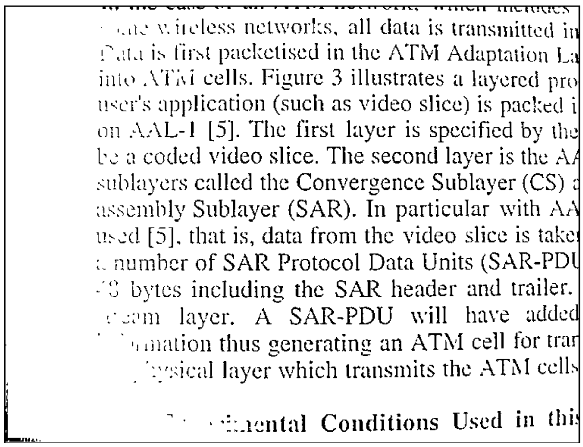
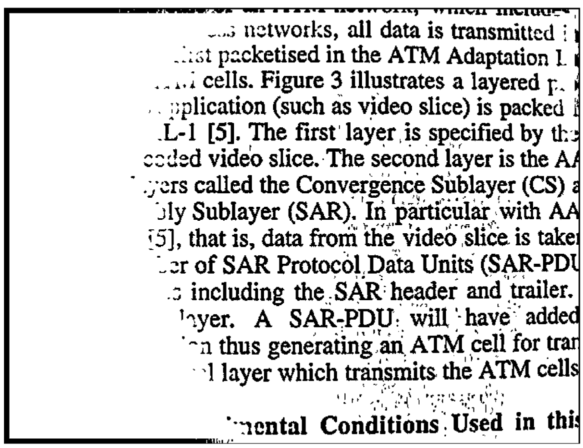

Several cell loss resilient techniques have been simulated so
emporal replacement; they are:
Simple resuming at the first slice header without motion
concealment technique (MCCT), after the lost cell.
Simple resuming at the first slice header with MCCT.

Fig. 27. Binary document image extracted using local contrast analysis from the image in Fig. 3.


wireless networks, all data is transmitted in
Data is first packetised in the ATM Adaptation La
into ATM cells. Figure 3 illustrates a layered pro
user's application (such as video slice) is packed i
on AAL-1 [5]. The first layer is specified by the
be a coded video slice. The second layer is the AA
sublayers called the Convergence Sublayer (CS) a
assembly Sublayer (SAR). In particular with AA
used [5], that is, data from the video slice is take
a number of SAR Protocol Data Units (SAR-PDU
bytes including the SAR header and trailer.
layer. A SAR-PDU will have added
mation thus generating an ATM cell for tra
sical layer which transmits the ATM cells
mental Conditions Used in thi

(a)

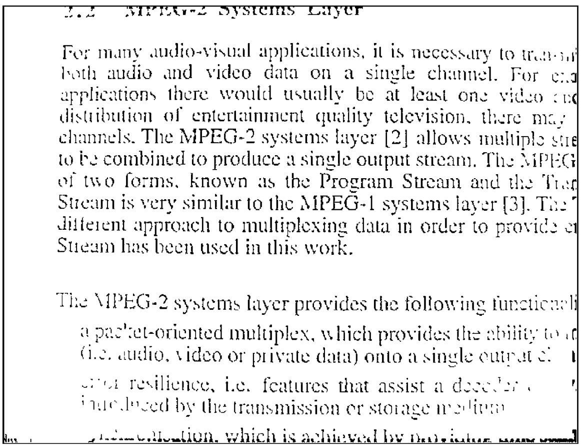

networks, all data is transmitted
packetised in the ATM Adaptation L
cells. Figure 3 illustrates a layered
pplication (such as video slice) is packed
L-1 [5]. The first layer is specified by
coded video slice. The second layer is the A
yers called the Convergence Sublayer (CS) a
bly Sublayer (SAR). In particular with AA
[5], that is, data from the video slice is take
of SAR Protocol Data Units (SAR-PDU
including the SAR header and trailer.
layer. A SAR-PDU will have added
n thus generating an ATM cell for tra
l layer which transmits the ATM cells
mental Conditions Used in thi

(b)

Fig. 28. Binary document image extracted using original logical level technique from the document image in Fig. 1. The results in (a) and (b) are quite different due to different predetermined parameters.

Table 1
Implementation results and evaluation. Execution time is the average processing time for several images of size 768 × 576 on a Sun Sparc Station IPX

Method	Average CUP time (s)	Subjective evaluation
Connectivity-based	46.583	Shadows
Local intensity gradient (post-processing)	51.166	Noise, unwanted edges
Integrated function	18.743	Noise
Local contrast	35.383	Noise, over-removal of shadow area
Logical level (fast algorithm [22])	12.166	Good, a little bit over-removal of shadow area
Adaptive logical level (fast algorithm [22] and postproc.)	15.533	Best


2.2 MPEG-2 Systems Layer

For many audio-visual applications, it is necessary to transmit both audio and video data on a single channel. For exa applications there would usually be at least one video and distribution of entertainment quality television, there may channels. The MPEG-2 systems layer [2] allows multiple stre to be combined to produce a single output stream. The MPEG of two forms, known as the Program Stream and the Tran Stream is very similar to the MPEG-1 systems layer [3]. The different approach to multiplexing data in order to provide e Stream has been used in this work.

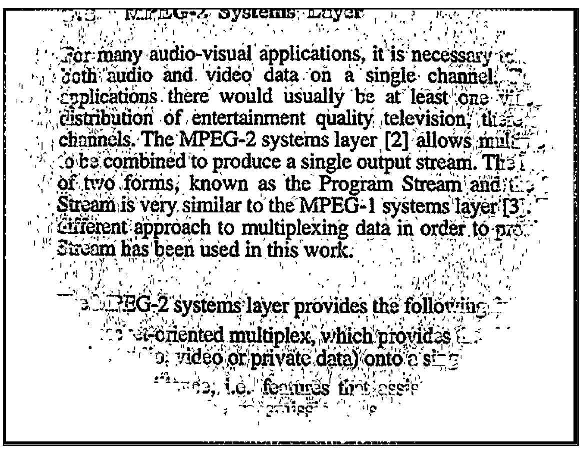
The MPEG-2 systems layer provides the following functionali

a packet-oriented multiplex, which provides the ability to (i.e. audio, video or private data) onto a single output cha

error resilience, i.e. features that assist a decoder introduced by the transmission or storage medium

synchronisation, which is achieved by providing

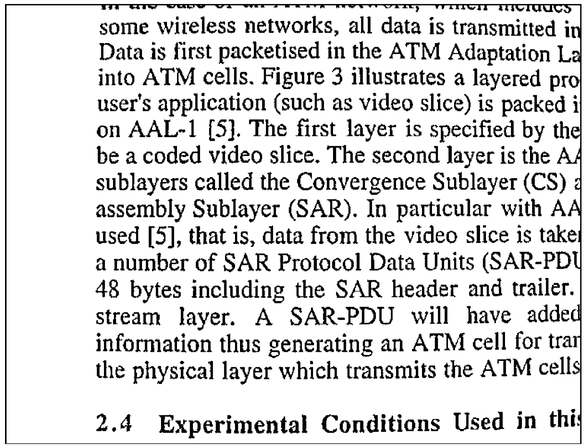
(a)


For many audio-visual applications, it is necessary to both audio and video data on a single channel. applications there would usually be at least one distribution of entertainment quality television, channels. The MPEG-2 systems layer [2] allows to be combined to produce a single output stream. of two forms, known as the Program Stream and Stream is very similar to the MPEG-1 systems layer [3]. different approach to multiplexing data in order to Stream has been used in this work.

MPEG-2 systems layer provides the following oriented multiplex, which provides video or private data) onto a features

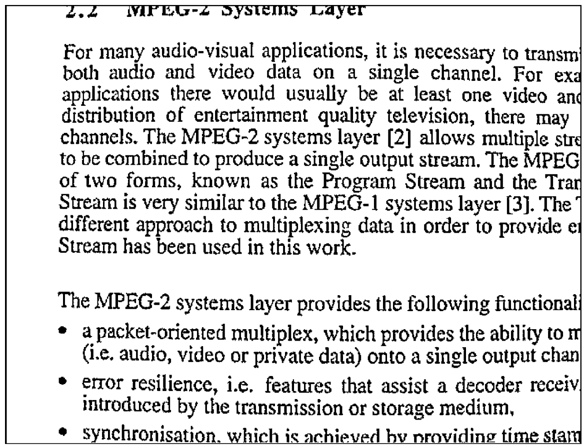
(b)

Fig. 29. Binary document image extracted using original logical level technique from the document image in Fig. 2. The results in (a) and (b) are quite different due to different predetermined parameters.


some wireless networks, all data is transmitted in Data is first packetised in the ATM Adaptation La into ATM cells. Figure 3 illustrates a layered pro user's application (such as video slice) is packed i on AAL-1 [5]. The first layer is specified by the be a coded video slice. The second layer is the A sublayers called the Convergence Sublayer (CS) assembly Sublayer (SAR). In particular with AA used [5], that is, data from the video slice is take a number of SAR Protocol Data Units (SAR-PDU 48 bytes including the SAR header and trailer. stream layer. A SAR-PDU will have added information thus generating an ATM cell for tra the physical layer which transmits the ATM cells

2.4 Experimental Conditions Used in thi

Fig. 30. Binary document image extracted using modified logical level thresholding method from the original Fig. 1. SW is taken as the run-length of the highest peak in the run-length histogram.

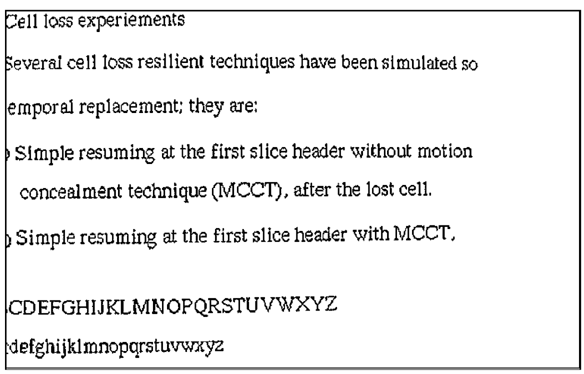

2.2 MPEG-2 Systems Layer

For many audio-visual applications, it is necessary to transmi both audio and video data on a single channel. For exa applications there would usually be at least one video and distribution of entertainment quality television, there may channels. The MPEG-2 systems layer [2] allows multiple stre to be combined to produce a single output stream. The MPEG of two forms, known as the Program Stream and the Tran Stream is very similar to the MPEG-1 systems layer [3]. The different approach to multiplexing data in order to provide e Stream has been used in this work.

The MPEG-2 systems layer provides the following functionali

- a packet-oriented multiplex, which provides the ability to m (i.e. audio, video or private data) onto a single output chan
- error resilience, i.e. features that assist a decoder receiv introduced by the transmission or storage medium,
- synchronisation, which is achieved by providing time stam

Fig. 31. Binary document image extracted using modified logical level thresholding method from the original in Fig. 2. SW is taken as the run-length of the highest peak in the run-length histogram.


Cell loss experiements

Several cell loss resilient techniques have been simulated so

emporal replacement; they are:

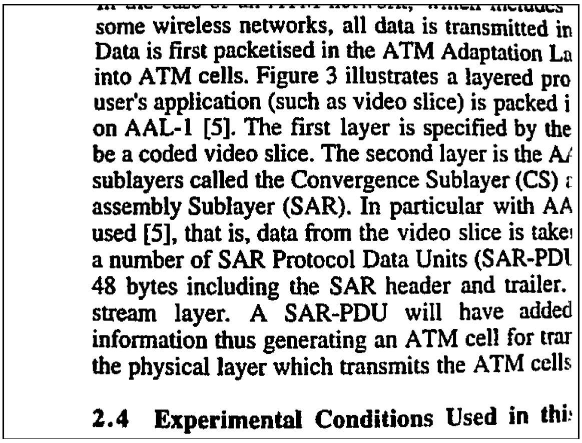
Simple resuming at the first slice header without motion concealment technique (MCCT), after the lost cell.

Simple resuming at the first slice header with MCCT.

CDEFGHIJKLMNOPQRSTUVWXYZ

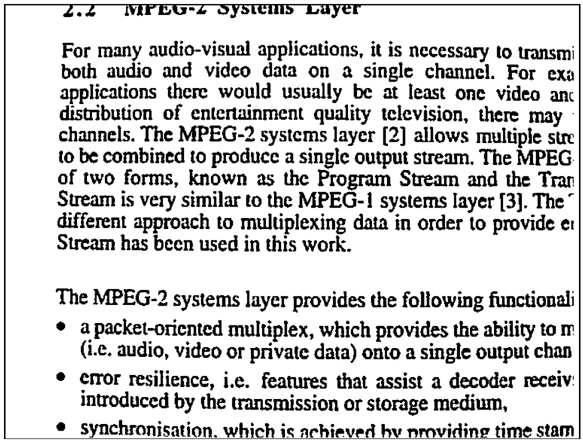
defghijklmnopqrstuvwxyz

Fig. 32. Binary document image extracted using modified logical level thresholding method from the original in Fig. 3. SW is taken as the run-length of the highest peak in the run-length histogram.


some wireless networks, all data is transmitted in
Data is first packetised in the ATM Adaptation La
into ATM cells. Figure 3 illustrates a layered pro
user's application (such as video slice) is packed i
on AAL-1 [5]. The first layer is specified by the
be a coded video slice. The second layer is the A/
sublayers called the Convergence Sublayer (CS) ε
assembly Sublayer (SAR). In particular with AA
used [5], that is, data from the video slice is take
a number of SAR Protocol Data Units (SAR-PDU
48 bytes including the SAR header and trailer.
stream layer. A SAR-PDU will have added
information thus generating an ATM cell for trar
the physical layer which transmits the ATM cells

2.4 Experimental Conditions Used in thi

Fig. 33. Binary document image extracted using modified logical level thresholding method from the original in Fig. 1. SW is taken as the run-length of the second highest peak right to the highest peak in the run-length histogram.


For many audio-visual applications, it is necessary to transmi
both audio and video data on a single channel. For exa
applications there would usually be at least one video anc
distribution of entertainment quality television, there may
channels. The MPEG-2 systems layer [2] allows multiple stre
to be combined to produce a single output stream. The MPEG
of two forms, known as the Program Stream and the Tran
Stream is very similar to the MPEG-1 systems layer [3]. The
different approach to multiplexing data in order to provide e
Stream has been used in this work.

The MPEG-2 systems layer provides the following functionali

- a packet-oriented multiplex, which provides the ability to m (i.e. audio, video or private data) onto a single output chan
- error resilience, i.e. features that assist a decoder receiv introduced by the transmission or storage medium,

Fig. 34. Binary document image extracted using our modified logical level thresholding method from the original in Fig. 2. SW is taken as the run-length of the second highest peak right to the highest peak in the run-length histogram.

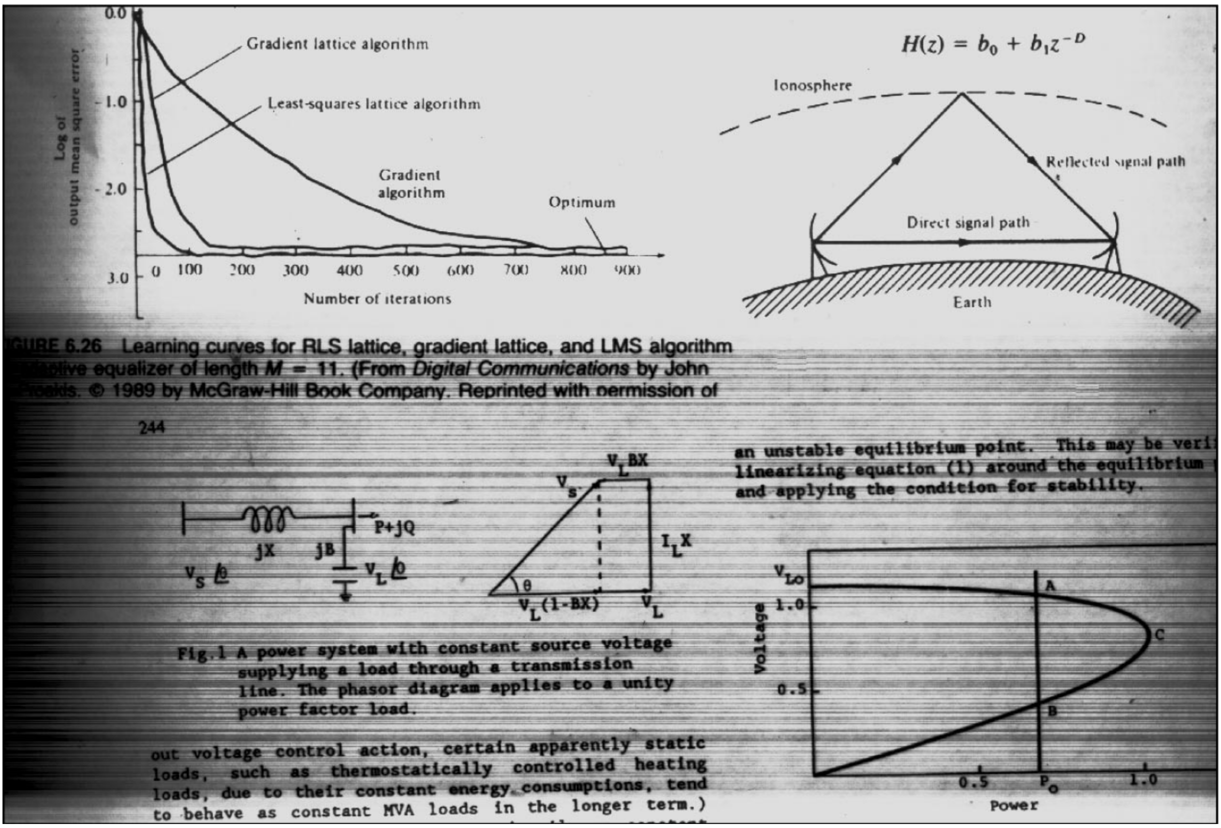


Fig. 35. An original gray-scale document image with a few of line graphics and shadows.

prints for noisy images and complex background images as shown in Figs. 20–24. The biggest difficulty for Local Contrast Technique is how to choose predetermined parameters, because it needs to set five parameters manually. Since only two out of three parameters T_3, T_4 and T_5 are independent, only T_1 and T_2 have clear physical meaning like in global thresholding or multi-threshold techniques and can be easily set. For a given image, it appears not to have some rules to set other parameters T_3, T_4 and T_5, and different parameters could produce quite different results. This method is also sensitive to inhomogeneous background and large shadows as shown in Figs. 25–27.

Logical Level Technique appears to work well for a wide range of document images, even though it also uses some *derivatives* in comparisons, its comparison is made for the local average and is not sensitive to noise. The result, however, could be changed from image to image as shown in Figs. 28 and 29 when the image contains some complex background or large change in illumination, because it uses a global predetermined parameter T. Our method improves its adaptivity and robustness from a predetermined global parameter to a local one. It can be implemented and tuned automatically from image to image, and is more insensitive to the local noises in images. The stroke width can be selected and adjusted automatically according to different document images and later pattern recognition requirements, that is, we can select the highest or second highest peak right to the highest peak of the run-length as stroke width under some conditions, therefore, it has a wider range of applications. Figs. 30–34 show the experimental results obtained using our method.

6. Conclusions

In this paper, we have presented a modified logical thresholding method based on adaptive logical level technique to binarize seriously degraded and very poor quality gray-scale document image. Our method can threshold gray-scale document images with complex signal-dependent noise, variable background intensity caused by nonuniform illumination, shadow, smear or smudge and very low contrast without obvious loss of useful information. It can adaptively tune the size of local analysing area and logical thresholding level according to the local run-length histogram for the selected regions

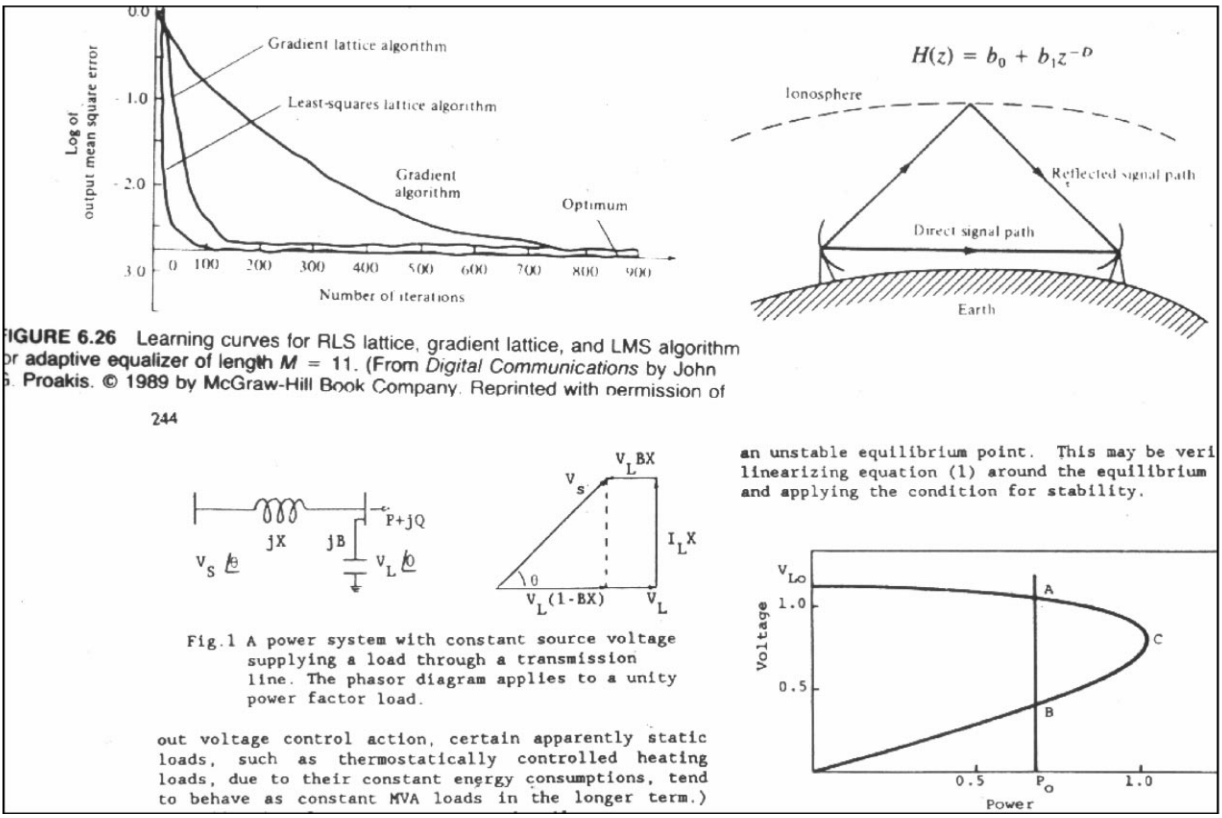
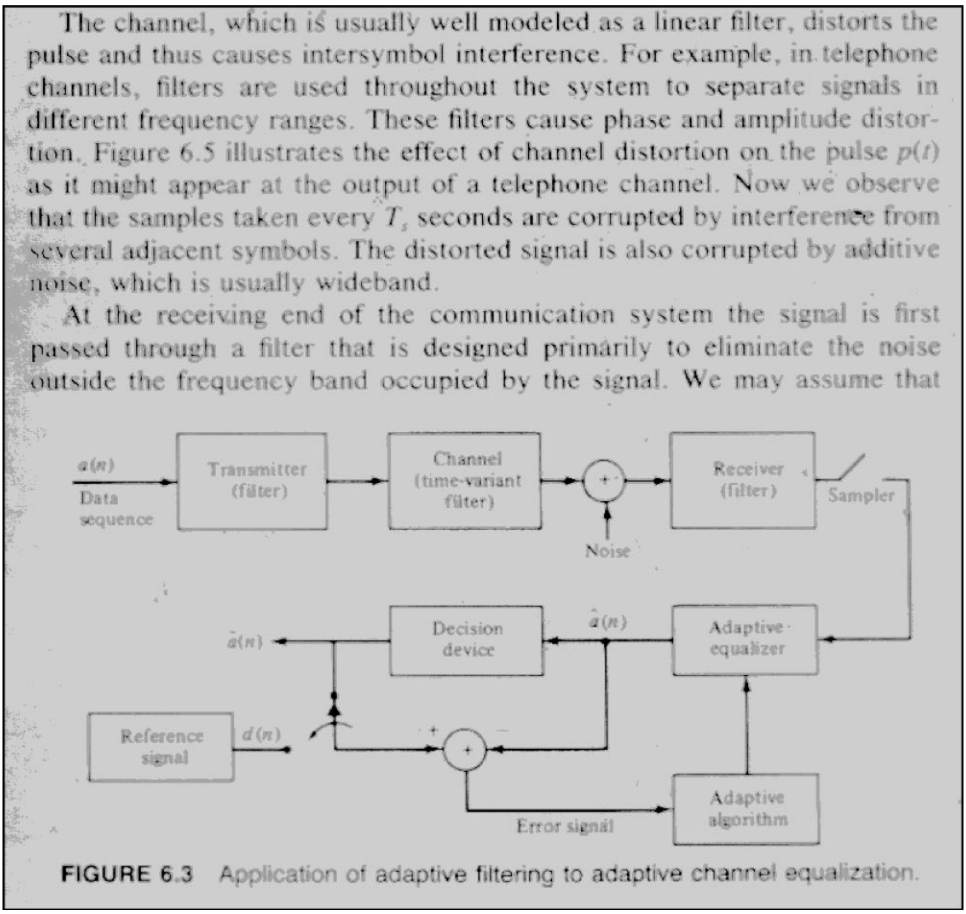


Fig. 36. Binary document image extracted using our modified logical level thresholding method from the original in Fig. 35.

with quasi-bimodal local histograms and the analysis of grayscale inhomogeneity of the background. For different test images with various noises and different inhomogeneous backgrounds, experiments and evaluations have shown that our method can automatically threshold various poor quality gray-scale document images without need of any prior knowledge of the document image and manual fine-tuning of parameters. It is nonparametric and automatic. It keeps useful information more accurately without overconnected and broken stroke of the characters, thus, it has the wider range of applications and is more robust for document images comparing with other thresholding methods based on connectivity and background analysis.

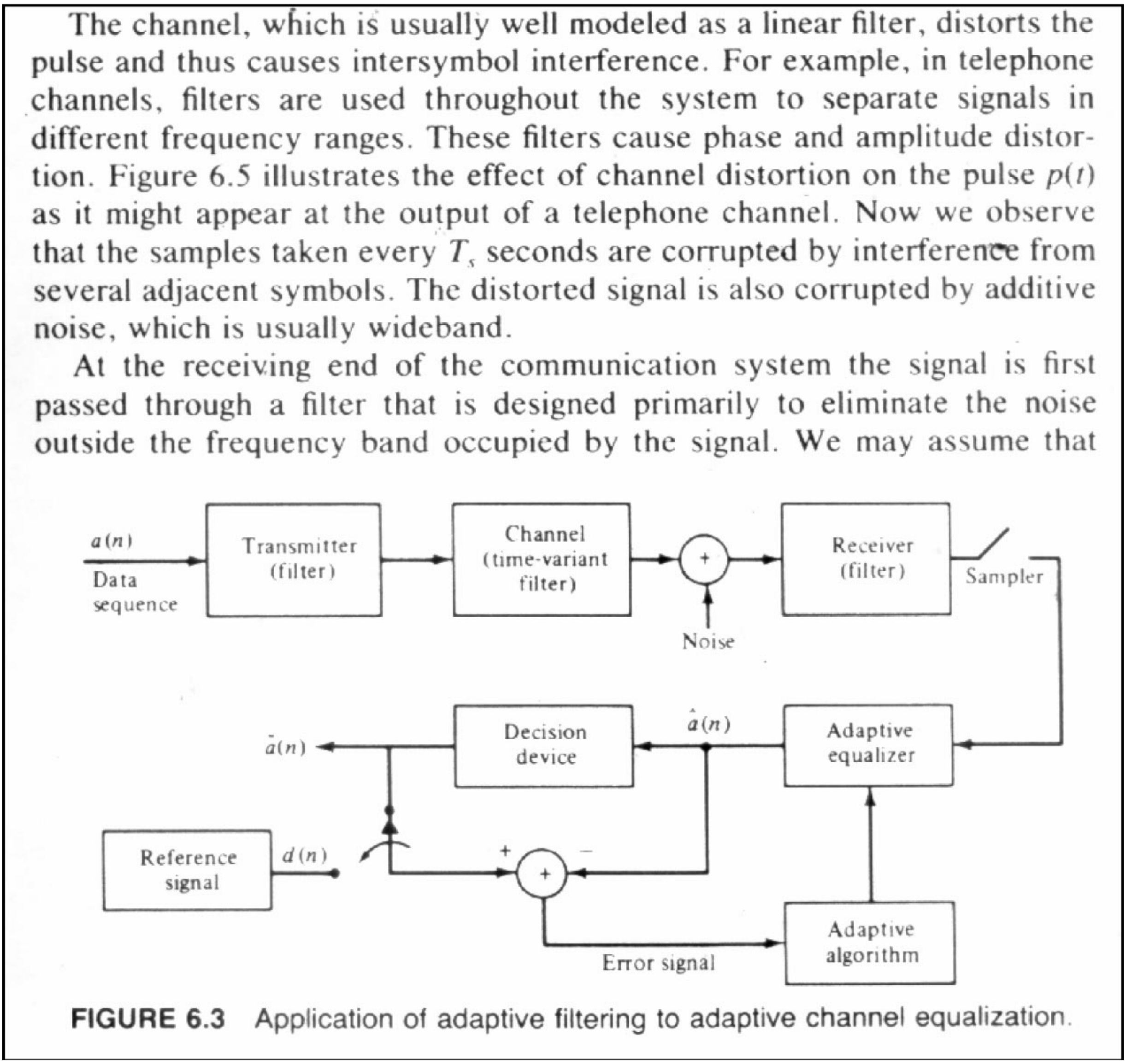
It is worth noting that our method is based on stroke width analysis, thus it can be used to process the document images with tables and line or block graphics and works well. Figs. 35 and 36 show an example of thresholding the document image with line graphics using our method. It may, however, not be suitable to threshold such gray-level images as scanned human or scenic photographies. Our method is a local adaptive technique, which is the modified logical level method. Its computation efficiency is much higher than connectivity-based thresholding method, as our method only needs to calculate a run-length histogram directly of the gray levels in the selected regions, instead of thresholding the whole original image at each intensity level to get its run-length histogram as in connectivity-based method. Experimental results show that user-defined parameter in our method is robust for various document images. Although our method is designed to process document images with very poor quality, it can perform equally well and work more efficiently on document images with good or normal quality because the background analysis of the document image, the run-length histogram construction and post-processing process are simpler. The average processing time for document images with good or normal quality can be reduced by 20–30%. Figs. 37 and 38 give an example of thresholding the document image with normal quality using our method.


The channel, which is usually well modeled as a linear filter, distorts the pulse and thus causes intersymbol interference. For example, in telephone channels, filters are used throughout the system to separate signals in different frequency ranges. These filters cause phase and amplitude distortion. Figure 6.5 illustrates the effect of channel distortion on the pulse $p(t)$ as it might appear at the output of a telephone channel. Now we observe that the samples taken every T_s seconds are corrupted by interference from several adjacent symbols. The distorted signal is also corrupted by additive noise, which is usually wideband.

At the receiving end of the communication system the signal is first passed through a filter that is designed primarily to eliminate the noise outside the frequency band occupied by the signal. We may assume that

FIGURE 6.3 Application of adaptive filtering to adaptive channel equalization.

Fig. 37. An original gray-scale document image with normal quality.


The channel, which is usually well modeled as a linear filter, distorts the pulse and thus causes intersymbol interference. For example, in telephone channels, filters are used throughout the system to separate signals in different frequency ranges. These filters cause phase and amplitude distortion. Figure 6.5 illustrates the effect of channel distortion on the pulse $p(t)$ as it might appear at the output of a telephone channel. Now we observe that the samples taken every T_s seconds are corrupted by interference from several adjacent symbols. The distorted signal is also corrupted by additive noise, which is usually wideband.

At the receiving end of the communication system the signal is first passed through a filter that is designed primarily to eliminate the noise outside the frequency band occupied by the signal. We may assume that

FIGURE 6.3 Application of adaptive filtering to adaptive channel equalization.

Fig. 38. Binary document image extracted using our modified logical level thresholding method from the original in Fig. 37.

Acknowledgements

This work is supported by the Australian Research Council.

References

[1] N. Ostu, A thresholding selection method from gray-level histogram, IEEE Trans. Systems Man Cybernet. SMC-8 (1978) 62–66.

[2] P.K. Sahoo, S. Soltani, A.K.C. Wong, A survey of thresholding technique, Comput. Vision Graphics Image Process. 41 (1988) 233–260.

[3] J.N. Kapur, P.K. Sahoo, A.K.C. Wong, A new method for gray-level picture thresholding using the entropy of the histogram, Computer Vision Graphics Image Process. 29 (1985) 273–285.

[4] S.U. Lee, S.Y. Chung, R.H. Park, A comparative performance study of several global thresholding techniques for segmentation, CVGIP 52 (1990) 171–190.

[5] F. Deravi, S.K. Pal, Gray level thresholding using second-order statistics, Pattern Recognition Lett. 1 (1983) 417–422.

[6] J. Kittler, J. Illingworth, Threshold selection based on a simple image statistic, CVGIP 30 (1985) 125–147.
[7] Y. Nakagawa, A. Rosenfeld, Some experiments on variable thresholding, Pattern Recognition 11 (1979) 191–204.
[8] S. Boukharouba, J.M. Rebordao, P.L. Wendel, An amplitude segmentation method based on the distribution function of an image, Computer Vision Graphics Image Process. 29 (1985) 47–59.
[9] S. Wang, R.M. Haralick, Automatic multithreshold selection, Computer Vision Graphics Image Process. 25 (1984) 46–67.
[10] R. Kohler, A segmentation system based on thresholding, Computer Graphics Image Process. 15 (1981) 319–338.
[11] N. Papamarkos, B. Gatos, A new approach for multilevel threshold selection, CVGIP: Graphical Models Image Process. 56 (5) (1994) 357–370.
[12] J.M. White, G.D. Rohrer, Imager segmentation for optical character recognition and other applications requiring character image extraction, IBM J. Res. Dev. 27 (4) (1983) 400–411.
[13] Y. Yasuda, M. Dubois, T.S. Huang, Data compression for check processing machine, Proc. IEEE 68 (7) (1980) 874–885.
[14] O.D. Trier, A.K. Jain, Goal-directed evaluation of binarization methods, IEEE Trans. Pattern Anal. Mach. Intell. 17 (12) (1995) 1191–1201.
[15] O.D. Trier, T. Taxt, Evaluation of binarization methods for document images, IEEE Trans. Pattern Anal. Mach. Intell. 17 (3) (1995) 312–315.
[16] J.S. Weszka, A. Rosenfeld, Threshold evaluation techniques, IEEE Trans, System Man Cybernet. SMC-8 (8) (1978) 622–629.
[17] Y. Liu, S.N. Srihari, Document image binarization based on texture features, IEEE Trans. Pattern Anal. Mach. Intell. 19 (5) (1997) 540–544.
[18] L. O'Gorman, Binarization and multithresholding of document images using connectivity, CVGIP: Graphical Models Image Process. 56 (6) (1994) 494–506.
[19] T. Taxt, P.J. Flynn, A.K. Jain, Segmentation of document image, IEEE Trans. Pattern Anal. Mach. Intell. 11 (12) (1989) 1322–1329.
[20] J.R. Parker, Gray level thresholding in badly illuminated images, IEEE Trans. Pattern Anal. Mach. Intell. 13 (8) (1991) 813–819.
[21] O.D. Trier, T. Taxt, Improvement of 'intergrated function algorithm' for binarization of document images, Pattern Recognition Lett. 16 (3) (1995) 277–283.
[22] M. Kamel, A. Zhao, Extraction of binary character/graphics images from grayscale document images, CVGIP: Graphical Models Image Process. 55 (3) (1993) 203–217.
[23] E. Giuliano, O. Paitra, L. Stringa, Electronic character reading system, U. S. Patent 4, 047,15, 6 September, 1977.

About the Author—YIBING YANG received her B.S., M.S. and Ph.D. degrees from Nanjing University of Aeronautics and Astronautics, China, in 1983, 1986 and 1991 respectively, all in electrical engineering. From 1986 to 1988, she worked as an assistant professor in Nanjing University of Aeronautics and Astronautics, China. From 1992 to 1993 she was a postdoctor, and from 1994, she has worked as an associate professor, both in the Department of Radio Engineering at Southeast University, China. Meanwhile, she was on leave and worked as a research associate in Electronics Department of The Chinese University of Hong Kong from 1995 to 1996. She is currently working as a visiting scholar in the Department of Electrical Engineering, The University of Sydney, Australia. Her research interests include image and signal analysis, processing and compression, pattern recognition, medical and optical image processing, and computer vision application.

About the Author—HONG YAN received his B.E. degree from Nanking Institute of Posts and Telecommunications in 1982, M.S.E. degree from the University of Michigan in 1984, and Ph.D. degree from Yale University in 1989, all in electrical engineering. From 1986 to 1989 he was a research scientist at General Network Corporation, New Haven, CT, USA, where he worked on developing a CAD system for optimizing telecommunication systems. Since 1989 he has been with the University of Sydney where he is currently a Professor in Electrical Engineering. His research interests include medical imaging, signal and image processing, neural networks and pattern recognition. He is an author or co-author of one book, and more than 200 technical papers in these areas. Dr. Yan is a fellow of the Institution of Engineers, Australia (IEAust), a senior member of the IEEE, and a member of the SPIE, the International Neural Network Society, the Pattern Recognition Society, and the International Society for Magnetic Resonance in Medicine.