Imaging Techniques in Document Analysis Processes

4

Basilis G. Gatos

Contents

B.G. Gatos
Institute of Informatics and Telecommunications, National Center for Scientific Research "Demokritos", Agia Paraskevi, Athens, Greece
e-mail: bgat@iit.demokritos.gr

D. Doermann, K. Tombre (eds.), *Handbook of Document Image Processing and Recognition*, DOI 10.1007/978-0-85729-859-1_4,

Abstract

Imaging techniques are widely used in document image analysis in order to process, enhance, analyze, and recognize document images. In this chapter, we present an overview of basic image processing algorithms used in document image analysis and focus on the techniques used for document image binarization, enhancement, and normalization.

Keywords

Binarization • Border removal • Color reduction • Dewarping • Image enhancement • Morphological operations • Page curl correction • Skew correction

Introduction

Several image processing algorithms are used at different stages of document image analysis process. These include morphological operations, skeletonization, connected component labeling, run-length smoothing, distance calculation, Hough transform, and projection profiles. In this chapter, we present an overview of these algorithms as well as an analysis of the imaging techniques used for document image binarization, enhancement, and normalization.

Document image binarization is used to separate the text from the background regions. It is an important, critical, and, at the same time, difficult and challenging task due to possible image degradations, nonuniform background intensity, low contrast, shadows, smear, etc. Document image enhancement aims to improve the quality of document images by diminishing artifacts such as low contrast and uneven background illumination, bleed-through and shadow effects, damaged characters, and noisy black borders. Document image normalization refers to the task of restoring the document image horizontal alignment after correcting possible page skew, character slant, warping, and perspective distortions.

The organization of this chapter is as follows. Section "Basic Image Processing Algorithms" presents the basic image processing algorithms used in document analysis process. Section "Document Image Binarization" gives an overview of document image binarization methods while section "Document Image Enhancement" presents document image enhancement state-of-the-art techniques. In section "Document Image Normalization," document image normalization methods are presented. Finally, a summary of the key issues of this chapter is given.

Basic Image Processing Algorithms

Digital image processing is a subcategory of digital signal processing and refers to the use of computer algorithms in order to process digital images. The development of image processing algorithms started in 1960s and till now has grown considerably with several applications including medical image processing, remote sensing, robot vision, image transmission, and coding. A detailed description of the available image processing algorithms can be found in several survey papers and books [55]. In this section, basic image processing algorithms used in document analysis process will be introduced. These include morphological operations, skeletonization, connected component labeling, run-length smoothing, distance calculation, Hough transform, and projection profiles.

Morphological Operations

Mathematical morphology was introduced in 1960s for describing the structure of materials by image analysis and is a very popular nonlinear theory for image processing, based on set theory. Originally developed for binary images, it was later generalized for grayscale images. The basic idea in binary morphology is to probe an image with a simple, predefined shape, drawing conclusions on how this shape fits or misses the shapes in the image. The two most basic operations in mathematical morphology are erosion and dilation. These operations work by translating a structuring element to various points in the input image and examining the intersection between the translated kernel coordinates and the input image coordinates. Morphological operators have been used extensively for various machine vision and recognition tasks. In document image analysis, morphological operations have been proved useful for several tasks like image cleaning and noise removal, layout analysis, skew correction, text line finding, and feature extraction.

Skeletonization

Skeletonization or thinning refers to the core-line detection of images. The purpose of skeletonization is to reduce the image components to their essential information so that further analysis and recognition are facilitated. In document image analysis, skeletonization is commonly used at the preprocessing, segmentation, and feature extraction stages.

Several methods for image skeletonization have been proposed in the literature and can be categorized in the following three broad categories. One category of skeletonization methods is based on distance transforms. Those methods detect all points that correspond to the centers of the maximal discs contained in the given image. A second category of methods produces median or center lines of the digital object in a non-iterative way. Usually, some critical image points are first calculated and then a specific path is defined by connecting these points. A third category of skeletonization methods is characterized by iterative thinning. In every iteration, every pixel is examined to be removed based on several pixel connectivity rules. The skeleton can be calculated using the basic operations of mathematical morphology which makes the skeleton a morphological representation.

Connected Component Labeling

Connected component labeling is used to assign each image region a unique label, enabling the distinct objects to be distinguished. In binary images, a label is assigned to each image pixel so that connected pixels have the same label. There are two common ways of defining connected pixels for a 2-D image: 4-connected pixels (connected only horizontally or vertically) and 8-connected pixels (connected horizontally, vertically, or diagonally). Connected component labeling is an important step in many image processing applications. In document image processing, it is used mainly at the preprocessing and the segmentation stages.

Connected component labeling algorithms are divided into three broad categories: multi-pass algorithms, two-pass algorithms, and one-pass algorithms.

Run-Length Smoothing Algorithm (RLSA)

In binary images, white runs correspond to successive horizontal or vertical background pixels. RLSA is one of the most popular document imaging techniques and is based on examining the white runs existing in the horizontal and vertical directions. For each direction, white runs whose lengths are smaller than a threshold smoothing value are eliminated [73]. RLSA is usually used for image enhancement and segmentation as well as for object recognition. In document image analysis process, RLSA has been proved very useful mainly for preprocessing, segmentation, and layout analysis.

Distance Transform

According to the distance transform, each pixel is labeled by the shortest distance from the boundary of the region within which it is contained. Distance transform can be used to obtain the thinned image by keeping only points of maximum local distance measures. This thinned image combined with the distance values can be used as a concise and descriptive representation of the original image from which the original image can be reconstructed. Distance transformed is used for a wide range of document image processing tasks including preprocessing, segmentation, and feature extraction.

There are two general approaches to calculate the distance transform. The first is the iterative approach according to which on each iteration boundaries are peeled from regions and set to the distance from the original boundary. According to the second approach, a fixed number of passes through the image is needed. Usually, two passes following different path directions are sufficient.

Hough Transform

The Hough transform has emerged in recent decades as a powerful method for many image processing and pattern recognition applications. It involves a transformation from the image coordinate plane to parameter space, and it is useful when the objective is to find lines or curves that fit groups of individual points on the image plane. The purpose of Hough transform is to find imperfect instances of objects within a certain class of shapes by a voting procedure. This voting procedure is carried out in the parameter space, from which object candidates are obtained as local maxima in a so-called accumulator space that is explicitly constructed by the algorithm for computing the Hough transform. In document image analysis, it is mainly used for document skew estimation, text line detection, and feature extraction.

A major drawback of its implementation in large images is its relatively low speed. For the acceleration of the Hough transform, one can select only several characteristic image points or connected component centers in order to calculate the Hough space.

Projection Profiles

Projection profiles are based on image profiling (projecting the image) in various directions. By calculating the local minima of horizontal and vertical projections, we can define several segments into which the image can be divided. Projection profiles are also used for document skew detection and feature extraction.

Document Image Binarization

Image binarization refers to the conversion of a grayscale or color image into a binary image. In document image processing, binarization is used to separate the text from the background regions by using a threshold selection technique in order to categorize all pixels as text or non-text. It is an important and critical stage in the document image analysis and recognition pipeline since it permits less image storage space, enhances the readability of text areas, and allows efficient and quick further processing for page segmentation and recognition. In the literature, several other terms are also used for document image binarization such as thresholding, text/background separation or segmentation, and background elimination. Binarization has been a subject of intense research in the field of document image processing during the last years. It is a challenging task due to several difficulties such as image degradations, nonuniform background intensity, low contrast, shadows, and smear.

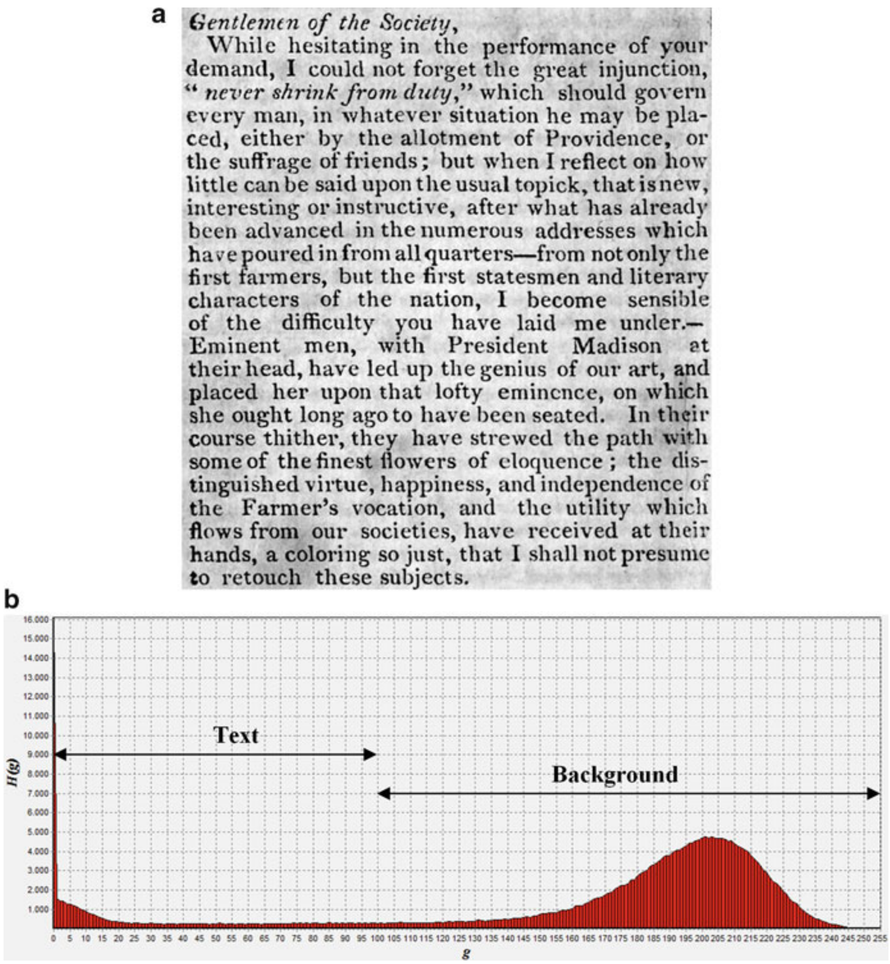
In general, document image binarization approaches are either global or local. In a global approach, threshold selection leads to a single threshold value T for the entire image. If $I(x, y)$ is the original grayscale image, then the resulting binary image $B(x, y)$ is defined as follows:

$$B(x, y) = \begin{cases} 1, & \text{if } I(x, y) \leq T \\ 0, & \text{if } I(x, y) > T \end{cases} \tag{4.1}$$

The grayscale histogram $H(g)$ of a document image that represents the distribution of the pixels in the image over the gray-level scale is presented in Fig. 4.1 (0 value corresponds to black while 255 to white). It can be observed that in this example, the histogram has a bimodal distribution: the left hill corresponds to text areas while the right hill to the background. In **global thresholding** techniques [16, 54, 61], the task is to calculate the optimal threshold T in order to separate these two hills. The selection of threshold T directly affects the quality of the produced binary image. As it is demonstrated in Fig. 4.2, a smaller T value may lead to broken or faint characters while a larger T value to noisy or merged characters in the resulting binary image.

Global thresholding has a good performance when there is a good separation between the foreground and the background areas (such as in Fig. 4.1b). If there is an overlap between these two areas, then global thresholding techniques will fail. An example of such a grayscale image is given in Fig. 4.3. In this example, if we apply a small global threshold $T = 130$ (see Fig. 4.4a), although background is almost completely removed, some text areas are also removed. A larger threshold ($T = 170$) leads to a binary result without missing any text information but with background areas that create a noisy image (see Fig. 4.4c). As we can also see in Fig. 4.4b, a threshold value between the above values ($T = 150$) leads to both missing text information and adding background noisy. In order to solve such problems, **local (adaptive) thresholding** techniques [27, 32, 50, 56, 77] have been introduced. According to these techniques, local area information guides the threshold value for each pixel of the image. These techniques have been widely used in document image analysis because they have a better performance in extracting the character strokes from an image that contains spatially uneven gray levels due to degradations.

There are also some **hybrid methods** [35, 68, 72] proposed for document image binarization that use both global and local information in order to decide if a pixel belongs to text or background category. Recently, it has been proposed [6, 28, 64] to **combine** and take into account the results of a set of binarization techniques in order to use the complementarity in success of each technique. Finally, a special category of document image binarization techniques is based on using **a training set and a machine learning** framework [14, 18, 30] while special techniques are used for the binarization of **color documents** [7, 67, 74] and are based on processing color information. In this section, the most representative works for


a

Gentlemen of the Society,
While hesitating in the performance of your demand, I could not forget the great injunction, "*never shrink from duty*," which should govern every man, in whatever situation he may be placed, either by the allotment of Providence, or the suffrage of friends; but when I reflect on how little can be said upon the usual topick, that is new, interesting or instructive, after what has already been advanced in the numerous addresses which have poured in from all quarters—from not only the first farmers, but the first statesmen and literary characters of the nation, I become sensible of the difficulty you have laid me under.— Eminent men, with President Madison at their head, have led up the genius of our art, and placed her upon that lofty eminence, on which she ought long ago to have been seated. In their course thither, they have strewed the path with some of the finest flowers of eloquence; the distinguished virtue, happiness, and independence of the Farmer's vocation, and the utility which flows from our societies, have received at their hands, a coloring so just, that I shall not presume to retouch these subjects.

Fig. 4.1 Grayscale histogram demonstration: (**a**) original grayscale image and (**b**) the corresponding grayscale histogram $H(g)$

the above mentioned document image binarization categories will be presented. Table 4.1 presents an overview of key document image binarization techniques. Comprehensive overviews of binarization techniques can be found in several survey papers [57, 66].

Global Thresholding Techniques

An efficient and frequently used global thresholding technique has been proposed by Otsu [54] and is based on histogram analysis. The threshold operation is regarded as the partitioning of the pixels of an image into two classes C_0 and C_1 where C_0

a

Gentlemen of the Society,

While hesitating in the performance of your demand, I could not forget the great injunction, "*never shrink from duty*," which should govern every man, in whatever situation he may be placed, either by the allotment of Providence, or the suffrage of friends; but when I reflect on how little can be said upon the usual topick, that is new, interesting or instructive, after what has already been advanced in the numerous addresses which have poured in from all quarters—from not only the first farmers, but the first statesmen and literary characters of the nation, I become sensible of the difficulty you have laid me under.—Eminent men, with President Madison at their head, have led up the genius of our art, and placed her upon that lofty eminence, on which she ought long ago to have been seated. In their course thither, they have strewed the path with some of the finest flowers of eloquence; the distinguished virtue, happiness, and independence of the Farmer's vocation, and the utility which flows from our societies, have received at their hands, a coloring so just, that I shall not presume to retouch these subjects.

b

Gentlemen of the Society,

While hesitating in the performance of your demand, I could not forget the great injunction, "*never shrink from duty*," which should govern every man, in whatever situation he may be placed, either by the allotment of Providence, or the suffrage of friends; but when I reflect on how little can be said upon the usual topick, that is new, interesting or instructive, after what has already been advanced in the numerous addresses which have poured in from all quarters—from not only the first farmers, but the first statesmen and literary characters of the nation, I become sensible of the difficulty you have laid me under.—Eminent men, with President Madison at their head, have led up the genius of our art, and placed her upon that lofty eminence, on which she ought long ago to have been seated. In their course thither, they have strewed the path with some of the finest flowers of eloquence; the distinguished virtue, happiness, and independence of the Farmer's vocation, and the utility which flows from our societies, have received at their hands, a coloring so just, that I shall not presume to retouch these subjects.

Fig. 4.2 (continued)

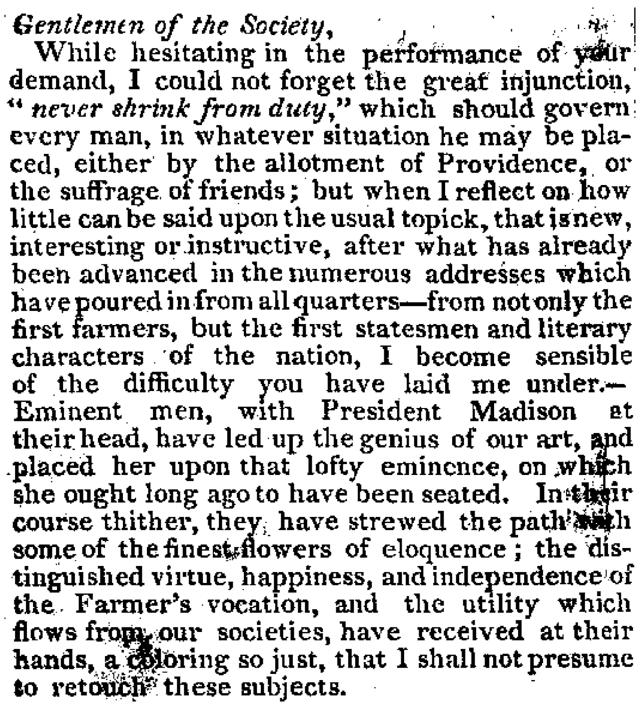
c 
Gentlemen of the Society,
While hesitating in the performance of your demand, I could not forget the great injunction, "*never shrink from duty*," which should govern every man, in whatever situation he may be placed, either by the allotment of Providence, or the suffrage of friends; but when I reflect on how little can be said upon the usual topick, that is new, interesting or instructive, after what has already been advanced in the numerous addresses which have poured in from all quarters—from not only the first farmers, but the first statesmen and literary characters of the nation, I become sensible of the difficulty you have laid me under.— Eminent men, with President Madison at their head, have led up the genius of our art, and placed her upon that lofty eminence, on which she ought long ago to have been seated. In their course thither, they have strewed the path with some of the finest flowers of eloquence; the distinguished virtue, happiness, and independence of the Farmer's vocation, and the utility which flows from our societies, have received at their hands, a coloring so just, that I shall not presume to retouch these subjects.

Fig. 4.2 Binarization results of the image in Fig. 4.1a: (**a**) with global threshold $T = 100$ (a value near the ideal threshold), (**b**) with $T = 30$, and (**c**) with $T = 140$

is the foreground (text) and C_1 is the background when global threshold t is used. In order to measure how good threshold t is, the discriminant criteria maximizing η are used where η is a separability measure defined as

$$\eta = \frac{\sigma_B^2}{\sigma_T^2} \tag{4.2}$$

where σ_B^2 and σ_T^2 are the between-class variance and the total variance, respectively, and

$$\sigma_T^2 = \sum_{i=0}^{255} (i - \mu_\mathrm{T})^2 p_i, \quad \mu_\mathrm{T} = \sum_{i=0}^{255} i\, p_i \tag{4.3}$$

$$\sigma_B^2 = \omega_0 \omega_1 (\mu_1 - \mu_0)^2 \tag{4.4}$$

p_i is the probability of occurrence of gray-level i and defined as

$$p_i = \frac{H(i)}{N} \tag{4.5}$$

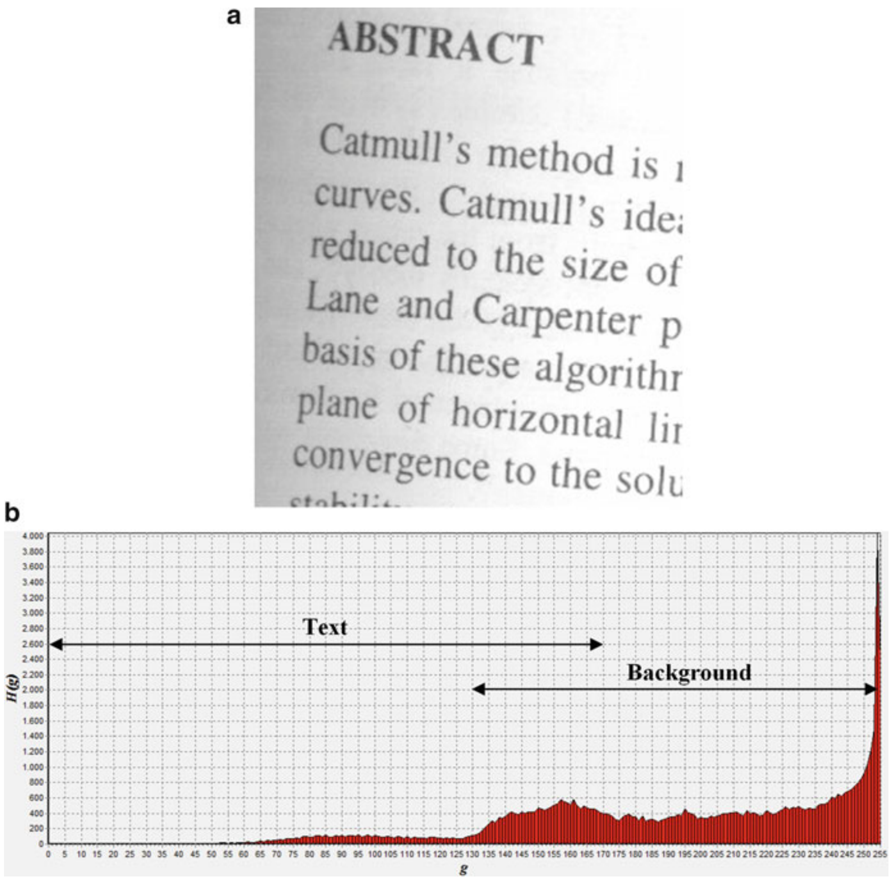


Fig. 4.3 A grayscale histogram of an image with overlapping text and background areas: (**a**) original grayscale image and (**b**) the corresponding grayscale histogram $H(g)$

where $H(i)$ is the grayscale histogram and N the total number of pixels. ω_0, ω_1, μ_0, and μ_1 are defined as follows:

$$\omega_0 = \sum_{i=0}^{t} p_i, \quad \omega_1 = 1 - \omega_0, \quad \mu_0 = \frac{\sum_{i=0}^{t} t\ p_i}{\omega_0}, \quad \mu_1 = \frac{\sum_{i=t+1}^{255} t\ p(i)}{\omega_1} \tag{4.6}$$

The optimal global threshold t_{opt} is defined as

$$t_{\text{opt}} = \underset{t \in [0..255]}{\arg\max}\, \eta \tag{4.7}$$

where η is defined by Eq. (4.2).

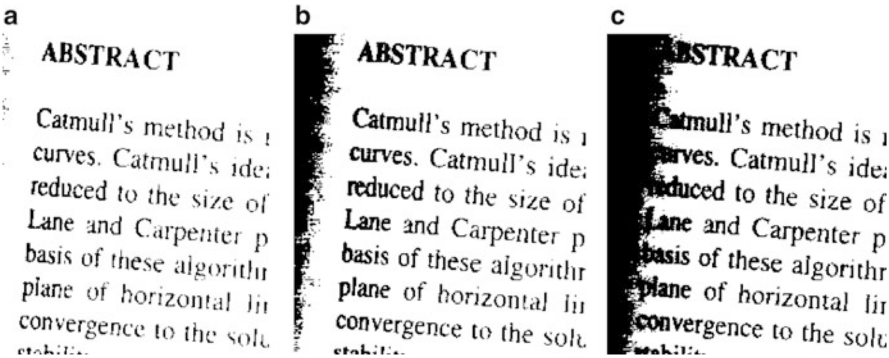


Fig. 4.4 Binarization results of the image in Fig. 4.3a: (**a**) with global threshold $T = 130$, (**b**) with $T = 150$, and (**c**) with $T = 170$

Otsu's algorithm can be successfully applied only to document images with a histogram that has a bimodal distribution. In order to treat document images with a non-uniform background, a modification of Otsu's approach has been proposed by Cheriet et al. [16] that is based on a recursive application of the above procedure. At each recursion, it segments first the object with the lowest intensity from the given image. The recursive process continues until there is only one object (the darkest) left in the image. It is recommended that the recursion stops when

$$\eta \geq 0.95 \tag{4.8}$$

where η is the separability measure defined by Eq. (4.2). The efficiency of this methodology has been demonstrated for the binarization of real-life bank checks.

According to iterative thresholding methods (see survey paper [57]), the optimum threshold is chosen automatically as a result of an iterative process. Successive iterations provide increasingly cleaner extractions of the text region. At iteration n, a new threshold T_n is calculated using the average of the foreground and background class means $m_f(T_n)$ and $m_b(T_n)$:

$$T_{n+1} = \frac{m_f(T_n) + m_b(T_n)}{2} \tag{4.9}$$

Iterations terminate when $|T_n - T_{n+1}|$ becomes sufficiently small.

Several binarization approaches exploit the entropy of the distribution of the gray levels and use the maximization of the entropy of the thresholded image as indicative of maximum information transfer (see survey paper [57]). Knowing the a priori entropy H_T of the gray-level histogram, the optimal threshold is estimated by maximizing the upper bound of the a posteriori entropy. The a priori entropy H_T is

Table 4.1 Overview of key document image binarization techniques

Reference	Category	Short description	Remarks
Otsu [54]	Global thresholding	The optimum threshold is calculated by separating the two classes of pixels (e.g., foreground and background) so that their between-class variance is maximized	Successful only to document images having bimodal distributed histogram
Cheriet et al. [16]	Global thresholding	A recursive application of [54] in order to treat document images with multiple background	Demonstrated for the binarization of bank checks
Solihin et al. [61]	Global thresholding	A two-stage thresholding approach requiring at a first stage each pixel to be assigned to one of three classes (foreground, background, fuzzy area)	Designed to binarize grayscale handwriting images
Niblack [50]	Local thresholding	A local binarization algorithm that calculates a pixelwise threshold by shifting a rectangular window across the image	Noise that is presented in the background may remain dominant in the final binary image
Sauvola and Pietikainen [56]	Local thresholding	A modification of [50] that adds a hypothesis on the gray values of text and background pixels	Noise is eliminated but text regions may be missed
Kamel and Zhao [32]	Local thresholding	The character stroke width and a global threshold are used in order to detect significant changes in gray-level values around the character body	It is very difficult to tune the parameters of the method
Yang and Yan [77]	Local thresholding	A modification of [32] in which character stroke width is automatically detected and the threshold is locally adapted	For poor quality grayscale document images without need of any manual parameter tuning
Gatos et al. [27]	Local thresholding	Adaptive method based on preprocessing, background surface calculation, suitable thresholding, and post-processing	For degraded documents with shadows, nonuniform illumination, low contrast, and noise
Kim et al. [35]	Hybrid thresholding	Global thresholding [54] on a difference image between the original and the water-filled image	Selection of two critical parameters on an experimental basis
Vonikakis et al. [72]	Hybrid thresholding	It combines the characteristics of the OFF center-surround cells of the Human Visual System with the classic Otsu binarization technique [54]	Experiments performed on a set of images with various computer-generated degradations
Tseng and Lee [68]	Hybrid thresholding	Based on analyzing the document layout and the intensity distribution. Pixels outside the detected blocks are binarized using Otsu technique [54]	Tested on a large set of images including newspapers, magazine articles, and business cards

(*continued*)

Table 4.1 (continued)

Reference	Category	Short description	Remarks
Badekas and Papamarkos [6]	Combination of techniques	Combines the results of several global and local binarization algorithms using a Kohonen self-organizing map neural network	Tested on a variety of degraded document images
Gatos et al. [28]	Combination of techniques	Combines the binarization results of several global and adaptive methodologies using a majority voting strategy and incorporates edge map information	For historical and degraded document images
Su et al. [64]	Combination of techniques	Image pixels are labeled as foreground, background, or uncertain based on the results of several binarization techniques. Uncertain pixels are classified at a next step as foreground or background	It improves the performance of existing binarization methods, robust on degraded images
Chou et al. [18]	Using training samples	Divides the image into regions and decides how to binarize each region using a learning process and a support vector machine (SVM) classifier	Better visual quality and OCR performance than several global and local methods
Huang et al. [30]	Using training samples	Hidden Markov model (HMM)-based binarization using feature vectors calculated from several image directions	Better OCR performance than other conventional methods
Chen and Leedham [14]	Using training samples	Local feature vectors are used to analyze and find the best approach to threshold a local area after quad-tree decomposition	Trained using 300 historical images and tested on 300 degraded document images
Badekas et al. [7]	For color documents	A color reduction procedure is followed by a connected component analysis in order to classify components as text or non-text	Appropriate for noisy color or grayscale documents
Tsai and Lee [67]	For color documents	The color image is transformed into the luminance and saturation spaces. A decision tree is used to decide if luminance, saturation, or both features will be employed for the binarization process	Trained on 150 different color documents and tested on an additional 519 color documents
Wang et al. [74]	For color documents	Color text image binarization based on color clustering and binary texture feature analysis. Two kinds of texture features, run-length histogram and spatial-size distribution related, respectively, are extracted and explored	Experiments with more than 500 color text images collected from the Internet

calculated as follows:

$$H_T = -\sum_{i=0}^{255} p_i \ln p_i \tag{4.10}$$

where p_i the probability of occurrence of gray-level I (see Eq. (4.5)).

Two probability distributions are assumed, one for the object area and one for the background. The a priori entropies for the two classes of pixels are defined as

$$H_b = -\sum_{i=1}^{t} \frac{p_i}{P_t} \ln\left(\frac{p_i}{P_t}\right), \quad H_w = -\sum_{i=t+1}^{255} \frac{p_i}{1-P_t} \ln\left(\frac{p_i}{1-P_t}\right) \tag{4.11}$$

where P_t is the cumulative probability function and calculated as follows:

$$P_t = \sum_{i=0}^{t} p_i \tag{4.12}$$

When the sum of the two class entropies H_b and H_w reaches its maximum, the image is considered as optimally thresholded.

A class of global thresholding methods designed to binarize grayscale handwriting images has been proposed by Solihin et al. [61]. It is based on a two-stage thresholding approach requiring at a first stage each pixel of handwritten image to be assigned to one of three classes: foreground, background, and a fuzzy area between them where it is hard to determine whether a pixel belongs to the foreground of the background. Parameters A and C that are used to define these three classes (see Fig. 4.5) are calculated based on an integral ratio function. At a second stage, the optimum threshold T was empirically calculated based on the handwriting media used:

$$T = \begin{cases} C - \frac{1}{2(C-A)}, & \text{if the writing device is a felt tipped pen} \\ C - \frac{1}{10(C-A)}, & \text{if the writing device is a ballpoint pen} \\ C, & \text{if the writing device is a pencil} \\ C - \frac{1}{10(C-A)}, & \text{if the writing device is not specified} \end{cases} \tag{4.13}$$

Local Thresholding Techniques

Niblack [50] introduced a local binarization algorithm that calculates a pixelwise threshold by shifting a rectangular window across the image. The threshold T for the center pixel of the window is computed using the mean m and the variance s of the gray values in the window:

$$T = m + ks \tag{4.14}$$

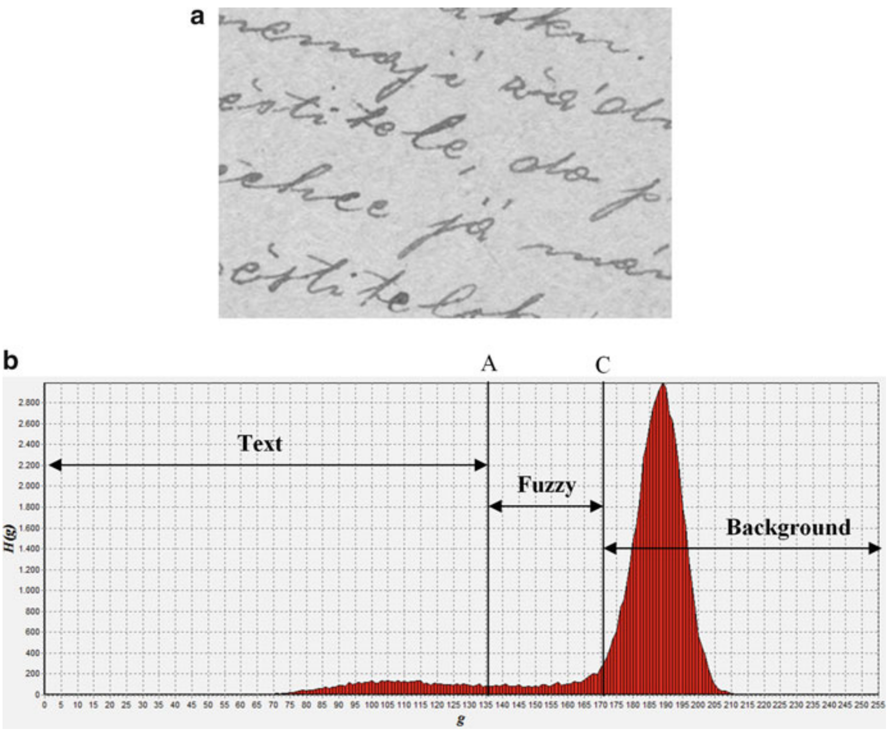


Fig. 4.5 A grayscale histogram of a handwriting image: (**a**) original grayscale image and (**b**) the corresponding three class grayscale histogram $H(g)$

where k is a constant set to -0.2. The value of k is used to determine how much of the total print object boundary is taken as a part of the given object. This method can distinguish the object from the background effectively in the areas close to the objects. The results are not very sensitive to the window size as long as the window covers at least one and two characters. However, noise that is present in the background remains dominant in the final binary image. Consequently, if the objects are sparse in an image, a lot of background noise will be left.

Sauvola and Pietikainen [56] propose a method that solves this problem by adding a hypothesis on the gray values of text and background pixels (text pixels have gray values near 0 and background pixels have gray values near 255), which results in the following formula for the threshold:

$$T = m + \left(1 - k\left(1 - \frac{s}{R}\right)\right) \tag{4.15}$$

where R is the dynamics of the standard deviation fixed to 128 and k takes on positive values (usually set to 0.5). Although this method gives better results for document images and noise is significantly eliminated, text regions may be also missed. Example results of applying local thresholding methods [50, 56] as well as global thresholding method [54] are presented in Fig. 4.6.

Logical level thresholding technique uses not only the image gray-level values but also the stroke width *SW* of the characters to improve the binarization quality. According to the algorithm of Kamel and Zhao [32], computations are performed according to *SW* and a global threshold which are predetermined by the user. In particular, the gray level or the smoothed gray level of each processing point is compared with four local averages located in $(2SW + 1)\times(2SW + 1)$ windows centered at two pairs of diametric points. This is expressed as follows:

$$b(x, y) = \begin{cases} 1, \ V_{i=0}^{3}[L(P_i) \wedge L(P_i^{'}) \wedge L(P_{i+1}) \wedge L(P_{i+1}^{'})] \text{ is true} \\ 0, \ \text{otherwise} \end{cases} \tag{4.16}$$

where $P_i^{'} = P_{(i+4) \bmod 8}$, $L(P) = \mathrm{ave}(P) - g(x, y) > T$, and

$$\mathrm{ave}(P) = \frac{\sum\limits_{-sw \le i \le sw} \sum\limits_{-sw \le j \le sw} f(P_x - i, P_y - j)}{(2sw + 1)^2} \tag{4.17}$$

where P_x and P_y are the coordinates of the point P and $g(x, y)$ is the gray level or its smoothed value. The original logical level technique of [32] is user-dependent since both the *SW* and the global threshold are predetermined by the user. Moreover, a global threshold is very difficult or even impossible to be tuned for document images with uneven illumination, big amount of noise, and other degradations. Yang and Yan [77] proposed the adaptive logical level technique (ALLT), in which the *SW* is automatically detected and the threshold is locally adapted. Concerning the *SW* detection in ALLT, the input image is divided in $N \times N$ $(N = 4, \ldots 8)$ regions with the aim of finding some local areas with quasi-bimodal histogram. Specifically, the histogram analysis is performed within regions of the two diagonal directions if N is even and additionally in the vertical and horizontal directions if N is odd. The quasi-bimodal regions are used for the run-length histogram analysis and the *SW* is defined as the run-length with the highest frequency. The other improvement proposed in ALLT concerns the local adaptive threshold. For each processing point P, the minimum (min), maximum (max), and average (ave) gray value are calculated within a $(2SW + 1)\times(2SW + 1)$ window centered at P. It is worth mentioning that if |max-ave| = |min-ave|, then the window expands to $(2SW + 3)\times(2SW + 3)$ and calculations are performed one more time. The adaptive threshold T is produced as follows:

$$T = \begin{cases} a\left(\frac{2}{3}\min + \frac{1}{3}\mathrm{ave}\right), \text{if } |\max - \mathrm{ave}| > |\min - \mathrm{ave}| \\ a\left(\frac{2}{3}\min + \frac{2}{3}\mathrm{ave}\right), \text{if } |\max - \mathrm{ave}| < |\min - \mathrm{ave}| \\ a \ \mathrm{ave}, \text{if } |\max - \mathrm{ave}| = |\min - \mathrm{ave}| \end{cases} \tag{4.18}$$

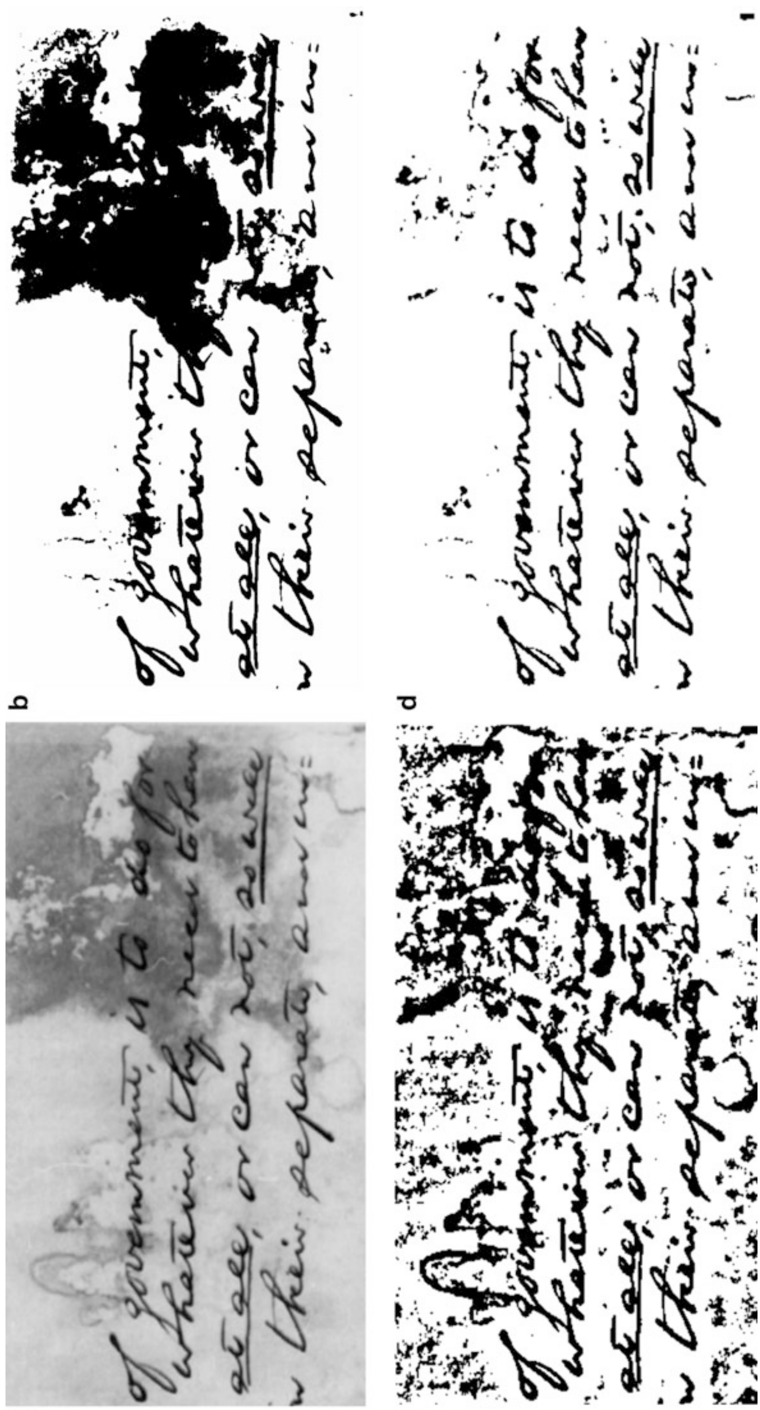


Fig. 4.6 Binarization results using global and local thresholding: **(a)** original grayscale image, **(b)** Otsu's method [54], **(c)** Niblack's method [50], and **(d)** Sauvola and Pietikainen method [56]

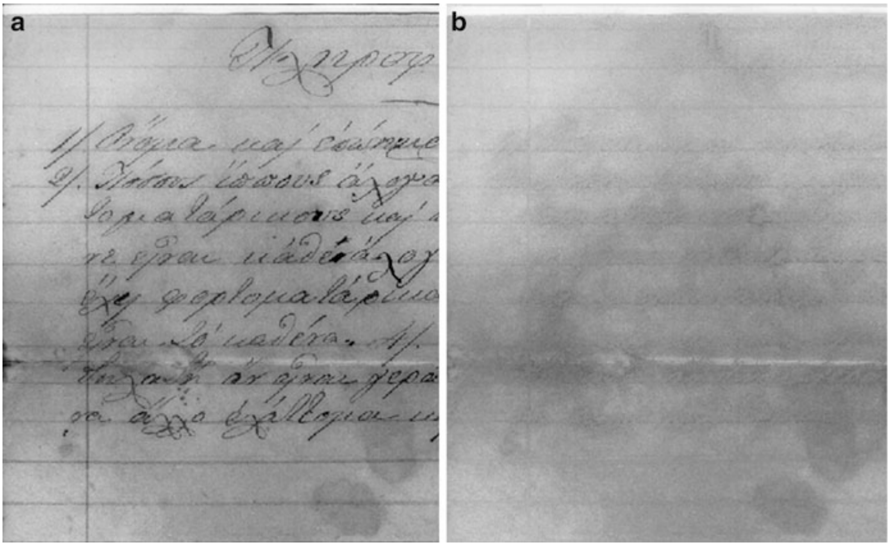


Fig. 4.7 An example of the background surface calculated according to the local thresholding (Gatos et al. [27]): (**a**) original grayscale image and (**b**) the calculated background surface

where α is a global predetermined parameter between 0.3 and 0.8, while 2/3 is recommended for most cases.

A local thresholding methodology has been proposed by Gatos et al. [27] for the binarization and enhancement of documents with degradations which occur due to shadows, nonuniform illumination, low contrast, large signal-dependent noise, smear, and strain. It follows several distinct steps: (a) a preprocessing procedure using a low-pass Wiener filter, (b) a rough estimation of foreground regions using Sauvola's approach [56] with a small k parameter (such as 0.1) in order to obtain a binary image that contains the foreground pixels plus some noise, (c) a background surface calculation by interpolating neighboring background intensities (see Fig. 4.7), (d) a thresholding by combining the calculated background surface with the original image using a threshold d that changes according to the grayscale value of the background surface B in order to preserve textual information even in very dark background areas, and, finally, (e) a post-processing step based on a consequent application of shrink and swell filtering in order to improve the quality of text regions and preserve stroke connectivity.

Hybrid Thresholding Techniques

Hybrid thresholding techniques first enhance the difference between the background and foreground and then binarize the enhanced image by a simple global algorithm.

Kim et al. [35] propose a local adaptive thresholding method where an image is regarded as a 3-D terrain and its local property is characterized by a water flow model. The water flow model locally detects the valleys corresponding to regions that are lower than neighboring regions. The deep valleys are filled with dropped water whereas the smooth plain regions keep up dry. The final step in this method concerns the application of a global thresholding such as Otsu's method [54] on a difference image between the original terrain and the water-filled terrain. A shortcoming of this method is the selection of two critical parameters for the method, namely, the amount of rainfall and the parameter of mask size, which is done on an experimental basis.

Vonikakis et al. [72] proposed a document binarization algorithm that combines the characteristics of the OFF center-surround cells of the Human Visual System with the classic Otsu binarization technique [54]. Cells of two different scales are combined, increasing the efficiency of the algorithm and reducing the extracted noise in the final output. A new response function, which regulates the output of the cell according to the local contrast and the local lighting conditions, is also introduced. The Otsu technique [54] is used to binarize the outputs of the OFF center-surround cells. Quantitative experiments performed on a set of images with various computer-generated degradations, such as noise, shadow, and low contrast, demonstrate the superior performance of the proposed method against six other well-established binarization techniques.

Tseng and Lee [68] proposed a document image binarization approach based on analyzing the document layout and the intensity distribution. All blocks with individual background intensity in a document image were first extracted by applying a two-stage block extraction technique. Then, the intensity distribution of each extracted block was analyzed to determine the ranges of background intensity. Those pixels in a block were binarized according to whether their intensity values are within the union range of background intensity. Other pixels outside all extracted blocks were binarized using Otsu's global thresholding method [54]. The proposed technique was tested on a large set of images that included newspapers, magazine articles, and business cards.

Combining Different Binarization Techniques

Recently, several binarization techniques combine the results of a set of binarization techniques in order to use the complementarity in success of each technique.

Badekas and Papamarkos [6] have proposed a document image binarization technique that takes advantage of the benefits of a set of selected global and local binarization algorithms by combining their results using a Kohonen self-organizing map neural network. Specifically, in the first stage the best parameter values for each independent binarization technique are estimated. In the second stage and in order to take advantage of the binarization information given by the independent techniques, the neural network is fed by the binarization results obtained by those techniques using their estimated best parameter values.

This procedure is adaptive because the estimation of the best parameter values depends on the content of images. The proposed technique is suitable to classify pixels that have high vagueness, that is, pixels which belong to edges and shadow areas and generally pixels that cannot be easily classified as foreground or background pixels. The proposed technique was tested on a variety of degraded document images. Motivated by the aforementioned technique, Gatos et al. [28] proposed a binarization methodology for historical and degraded document images that is based on (i) efficient preprocessing using the Wiener filter, (ii) combination of the binarization result of several global and adaptive state-of-the-art methodologies using a majority voting strategy, (iii) incorporation of the edge map in the grayscale image produced by the Canny edge detector, and (iv) application of efficient image post-processing based on mathematical morphology for the enhancement of the final result.

Su et al. [64] claim that by combining different binarization techniques, better performance can be achieved with careful analysis. Those pixels that are labeled the same by different methods are usually correctly classified, and those pixels which are classified as text by some methods and labeled as background by other methods have higher possibility to be misclassified than others. Based on such observation, we divide all the image pixels into three sets: foreground set, where those pixels are classified into foreground by all the examining binarization methods; background set, where those pixels are classified into background by all the examining binarization methods; and uncertain set, where the rest of pixels belong to, which is defined as follows:

$$P(x,y) = \begin{cases} \text{foreground}, & \sum_{i=1}^{n} B_i(x,y) = 0 \\ \text{background}, & \sum_{i=1}^{n} B_i(x,y) = n \\ \text{uncertain}, & \text{otherwise} \end{cases} \tag{4.19}$$

where $P(x,y)$ denotes one image pixel, and $B_i(x)$, which is either 0 (foreground) or 1 (background), denotes the corresponding binarization result of pixels $P(x,y)$ generated by the ith binarization methods. The pixels are then projected into a feature space. Those pixels in foreground and background sets can be viewed as correctly labeled samples and used to determine the label of those uncertain pixels. A classifier is then applied to iteratively classify those uncertain pixels into foreground and background.

Using Training Samples

A training procedure for binarization is used by Chou et al. [18] for document images produced by cameras. This method divides an image into several regions and decides how to binarize each region. The decision rules are derived from a learning process that takes training images as input. Within each region r, one of the following four operations is applied: set the whole of r to black, set the whole of r to white, use Otsu's method [54] to compute the threshold for r, or use the smallest Otsu threshold in the neighboring regions as the threshold for r. A learning process is used to establish the rules for deciding which of the above options should be adopted for each region. The rules are expressed as decision functions, which take a number of features extracted from r as input. For the learning process, a support vector machine (SVM) classifier is used having as input the following three features: $T_{\text{Otsu}}(r) - T_{\min}(r)$, $\mu(r)$, and $\sigma(r)$ where $T_{\text{Otsu}}(r)$ is the Otsu threshold for r, $\mu(r)$ and $\sigma(r)$ the mean and the standard deviation of the distribution of gray values in r, and $T_{\min}(r)$ is defined as follows:

$$T_{\min}(r) = \min\{T_{\text{Otsu}}(r), \min_{s \in \Lambda(r)} T_{\text{Otsu}}(s)\} \tag{4.20}$$

where $\Lambda(r)$ is the set of neighboring regions of r. Tests on images produced under normal and inadequate illumination conditions show that this method yields better visual quality and better OCR performance than three global binarization methods and four locally adaptive binarization methods.

In [30], a hidden Markov model (HMM)-based document image binarization algorithm is presented by Huang et al. In a first stage, a coarse global thresholding method is used to discriminate the bright part of the whole image from the foreground pixels which have lower values. The mean gray-level value is regarded as the preliminary threshold value of this process. In a second stage, the remaining pixels which are supposed to be a mixture of foreground and background are applied to the HMM pixel classifier to obtain the attribute of each pixel. A novel feature extraction method suitable for HMM-based image binarization is used. Inside a window of N by N, four feature vectors are calculated from the four following directions, horizontal, left up to right down, vertical, and right up to left down, around each of which contains N elements.

In [14], Chen and Leedham propose a machine learning framework for thresholding degraded historical document images. A new thresholding structure called the decompose algorithm is proposed and compared against some existing single-stage algorithms. The decompose algorithm uses local feature vectors to analyze and find the best approach to threshold a local area. Instead of employing a single thresholding algorithm, automatic selection of an appropriate algorithm for specific types of subregions of the document is performed. The original image is recursively broken down into subregions using quad-tree decomposition until a suitable thresholding method can be applied to each subregion. The algorithm has been trained using 300 historical images and evaluated on 300 "difficult" document images, in which considerable background noise or variation in contrast and illumination exists. Quantitative analysis of the results by measuring text recall and qualitative assessment of processed document image quality are reported. The decompose algorithm is demonstrated to be effective at resolving the problem in varying quality historical images.

Binarization of Color Documents

A method for the binarization of color document images is presented in [7] by Badekas et al. Initially, the colors of the document image are reduced to a small number using a new color reduction technique. Specifically, this technique estimates the dominant colors and then assigns the original image colors to them in order that the background and text components become uniform. Each dominant color defines a color plane in which the connected components (CCs) are extracted. Next, in each color plane, a CC filtering procedure is applied which is followed by a grouping procedure. At the end of this stage, blocks of CCs are constructed which are next redefined by obtaining the direction of connection (DOC) property for each CC. Using the DOC property, the blocks of CCs are classified as text or non-text. The identified text blocks are binarized properly using suitable binarization techniques, considering the rest of the pixels as background. The final result is a binary image which contains always black characters in white background independently of the original colors of each text block. The proposed document binarization approach can also be used for binarization of noisy color (or grayscale) document images.

According to the methodology proposed in [67] by Tsai and Lee, the color document image is initially transformed into the luminance and saturation spaces. Then, the luminance histogram is computed and a Gaussian smoothing filter is employed to remove unreliable peaks and valleys upon the histogram. The threshold candidates are then selected from the histogram. Next, the entire image distribution is analyzed to extract statistical features to be applied in a decision tree. The tree decides if luminance, saturation, or both features will be employed to binarize color document images. First, if the document image colors are concentrated within a limited range, saturation is employed. Second, if the image foreground colors are significant, luminance is adopted. Third, if the image background colors are concentrated within a limited range, luminance is also applied. Fourth, if the total number of pixels with low luminance (less than 60) is limited, saturation is applied; else, both luminance and saturation are employed. Finally, the color features are used for image binarization. The method was trained on 150 different color documents and tested on an additional 519 color documents. The performance analysis indicated that the method is efficient and effective in color documents with foreground and background colors either close or mixed.

A color text image binarization based on binary texture analysis is proposed in [74] by Wang et al. It is designed to overcome the limitations of existing techniques for color text images. This algorithm efficiently integrates color clustering and binary texture feature analysis. Two kinds of features capable of effectively characterizing text-like binary textures, including run-length histogram and spatial-size distribution-related features, are extracted and explored. In addition, in order to handle varying colors of the text in the image, a combination strategy is implemented among the binary images obtained by color clustering. The effective cooperation of these techniques enables the algorithm to survive those complex background conditions, as existing techniques tend to fail.

Document Image Enhancement

Document images usually suffer from several defects mainly due to natural ageing, environmental conditions, usage, poor storage conditions of the original, as well as human manipulations during the scanning process. Since document image analysis and recognition methods usually assume a smooth background and a good quality of printing or writing, it is imperative to have an efficient image enhancement process in order to restore the good quality of document images. Moreover, document image enhancement process enhances the readability of text areas and permits less image storage space. The main artifacts encountered in document images can be categorized into the following four broad categories:

Low contrast and uneven background illumination. At low contrast documents, text regions are not clear due to poor contrast between the foreground text and the background. Documents with uneven background illumination have a variable background intensity (see Fig. 4.8).

Bleed-through, shining, or shadow-through effects. Bleed-through is the result of the diffusion of ink from one side of a page through the other side. The letters from one page also appear on the other page when the printing ink was not dry or due to the long time of pressing two leaves together (see Fig. 4.9). A similar degradation phenomenon is the shining or shadow-through effect. In this case, if a paper is relatively thin, then because of the transparency of the paper the information on one side of document interferes with the information contained on the other side.

Damaged characters or noisy background. The characters may be of poor quality, fainted, and merged, with holes or noisy edges. Spots, smears, shadows, or noise may appear at the background (see Fig. 4.10).

Borders or parts of neighboring page. When capturing a document, the resulting document image is often framed by a noisy black border or includes noisy text regions from neighboring pages (see Fig. 4.11).

In this section, the main document image enhancement approaches that have been proposed in the literature to face all the above artifacts will be presented. An overview of the key document image enhancement techniques is given in Table 4.2.

Low Contrast and Uneven Background Illumination

A preprocessing stage of the document image is essential for contrast enhancement between background and text areas, smoothing of background texture, as well as the elimination of noisy areas. The use of a low-pass Wiener filter has been proved efficient for the above goals and has been used as preprocessing stage at a binarization process [27]. Contrast enhancement is a major issue in image processing, while histogram equalization is the most common, simple, and effective contrast

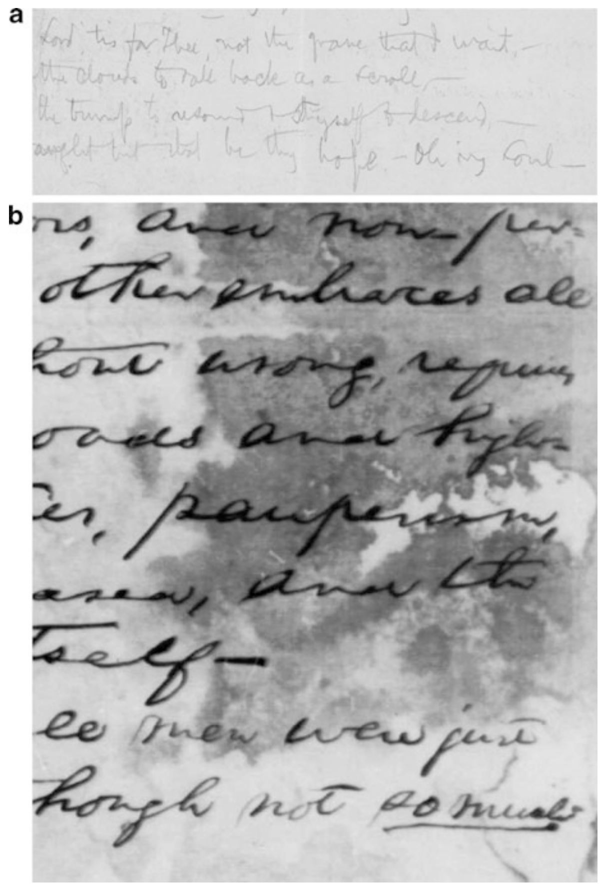
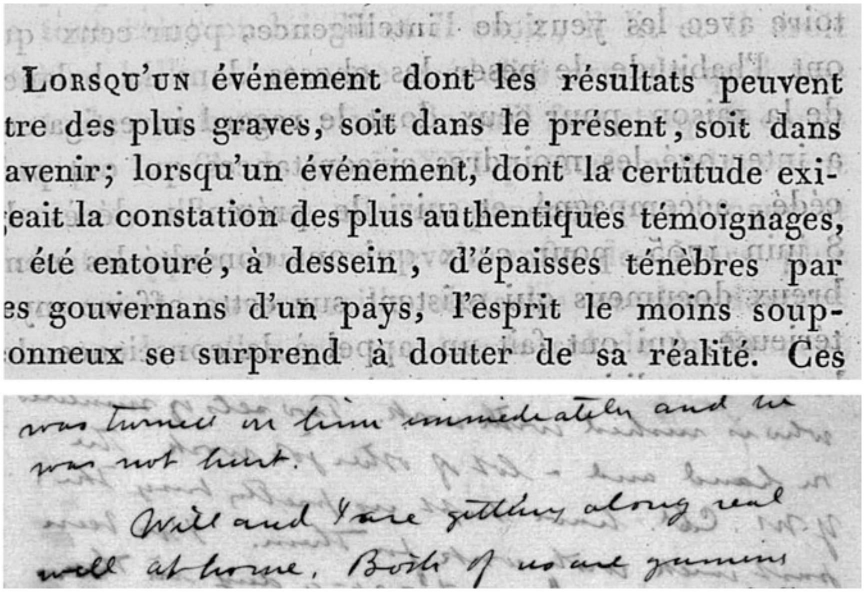


Fig. 4.8 Document image examples having (**a**) low contrast and (**b**) uneven background illumination

enhancement technique. A generalized fuzzy operator preprocessed with histogram equalization and partially overlapped sub-block histogram equalization is used from Leung et al. [40] for reducing background noise and increasing the readability of text by contrast enhancement. This method achieves good performance in contrast enhancement of extremely low contrast and low illuminated document images.

Total variation regularization flattens background gray levels and produces an image where background noise is considerably reduced. Nonlocal means filtering can smooth character parts and improve character quality based on neighboring data information. The above two techniques are combined by Likforman et al. [43] in order to enhance the quality of historical printed document images.

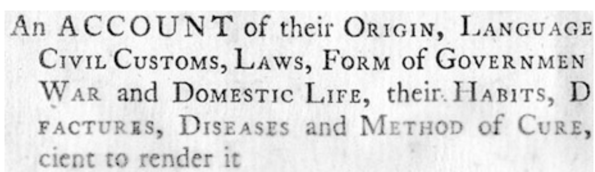
Morphological operations can be also used in order to enhance the background of document images. In Nomura et al. [51], morphological operations are used to remove undesirable shapes called critical shadows on the background of document images before proceeding to binarization. These critical shadows may appear due to changes in color or size, very low contrast, low quality of focalization, and poor, a b



Lorsqu'un événement dont les résultats peuvent
tre des plus graves, soit dans le présent, soit dans
avenir; lorsqu'un événement, dont la certitude exi-
;eait la constation des plus authentiques témoignages,
été entouré, à dessein, d'épaisses ténèbres par
es gouvernans d'un pays, l'esprit le moins soup-
onneux se surprend à douter de sa réalité. Ces

was turned in time immediately and he
was not hurt.
Will and I are getting along real
well at home. Both of us are gunning

Fig. 4.9 Parts of machine-printed and handwritten document images suffering (**a**) from the bleed-through and (**b**) from the shining-through effect



An ACCOUNT of their Origin, Language
Civil Customs, Laws, Form of Governmen
War and Domestic Life, their Habits, D
factures, Diseases and Method of Cure,
cient to render it

Fig. 4.10 An example of a document image with damaged characters and noisy background

and nonuniform illumination. Method [51] adaptively and without manual fine-tuning of parameters locates these critical shadows on grayscale degraded document images using morphological operations and lightens them before applying eventual thresholding process.

Bleed-Through, Shining, or Shadow-Through Effects

The methods proposed in the literature dealing with bleed-through, shining, or shadow-through effects can be categorized into either non-registration (or blind) or registration (or non-blind) methods, depending on whether the verso image of a page is available and precisely registered to the recto image.

Non-registration (Blind)-Based Methods

In non-registration-based methods, the two sides of a page are treated independently and processed separately. Those methods attempt to clean the front side of a

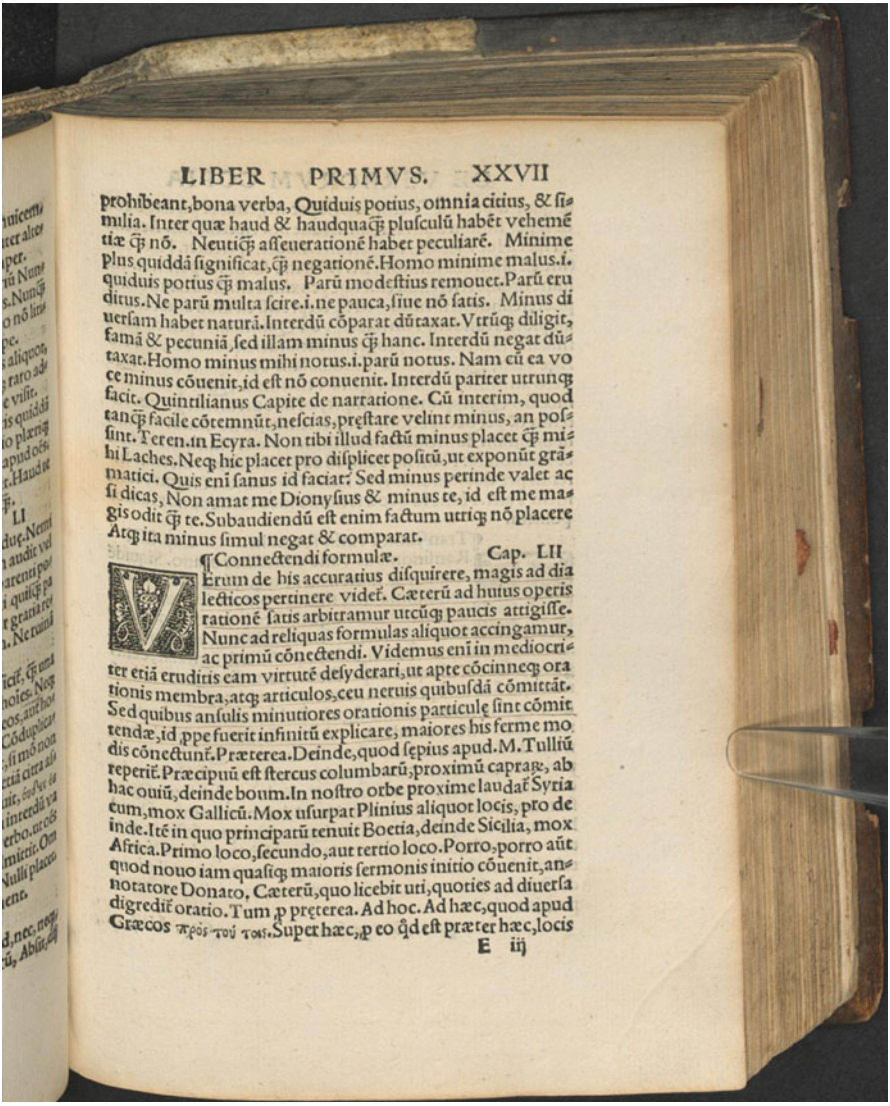

LIBER PRIMVS. XXVII
prohibeant,bona verba, Quiduis potius, omnia citius, & si-
milia. Inter quæ haud & haudquaq͂ plusculũ habẽt veheme
tiæ q͂ nõ. Neutiq͂ affeuerationẽ habet peculiarẽ. Minime
plus quiddã significat,q͂ negationẽ.Homo minime malus.i.
quiduis potius q͂ malus. Parũ modeftius remouet.Parũ eru
ditus.Ne parũ multa fcire.i.ne pauca,fiue nõ fatis. Minus di
uerfam habet naturã.Interdũ cõparat dũtaxat.Vtrũq; diligit,
famã & pecuniã,fed illam minus q͂ hanc. Interdũ negat dũ-
taxat.Homo minus mihi notus.i.parũ notus. Nam cũ ea vo
ce minus cõuenit,id eft nõ conuenit. Interdũ pariter utrunq;
facit. Quintilianus Capite de narratione. Cũ interim, quod
tanq͂ facile cõtemnũt,nefcias,preftare velint minus, an pof-
fint. Teren.in Ecyra. Non tibi illud factũ minus placet q͂ mi-
hi Laches.Neq; hic placet pro difplicet pofitũ,ut exponũt grã-
matici. Quis eni fanus id faciat? Sed minus perinde valet ac
fi dicas, Non amat me Dionyfius & minus te, id eft me ma-
gis odit q͂ te.Subaudiendũ eft enim factum utriq; nõ placere
Atq; ita minus fimul negat & comparat.
¶Connectendi formulæ. Cap. LII
VErum de his accuratius difquirere, magis ad dia
lecticos pertinere videt̃. Cæterũ ad huius operis
rationẽ fatis arbitramur utcũq; paucis attigiffe.
Nunc ad reliquas formulas aliquot accingamur,
ac primũ cõnectendi. Videmus enĩ in mediocri-
ter etiã eruditis eam virtutẽ defyderari,ut apte cõcinneq; ora
tionis membra,atq; articulos,ceu neruis quibufdã cõmittãt.
Sed quibus anfulis minutiores orationis particulę fint cõmit
tendæ,id ꝓpe fuerit infinitũ explicare, maiores his ferme mo
dis cõnectunt̃.Præterea.Deinde,quod fępius apud.M.Tulliũ
reperit̃.Præcipuũ eft ftercus columbarũ,proximũ capraȩ, ab
hac ouiũ,deinde boum.In noftro orbe proxime laudat̃ Syria
cum,mox Gallicũ.Mox ufurpat Plinius aliquot locis, pro de
inde.Itẽ in quo principatũ tenuit Boetia,deinde Sicilia, mox
Africa.Primo loco,fecundo,aut tertio loco.Porro,porro aũt
quod nouo iam quafiq; maioris fermonis initio cõuenit,an-
notatore Donato. Cæterũ,quo licebit uti,quoties ad diuerfa
digredit̃ oratio.Tum ꝓ pręterea. Ad hoc. Ad hæc,quod apud
Græcos πρὸς τοῦ τοῖς.Super hæc,ꝓ eo q̃d eft præter hæc,locis
E iij

Fig. 4.11 An example of a document image with noisy black border as well as with text part from the neighboring page

document without referring to the reverse side. Most of these methods treat bleed-through interference as a kind of background artifacts or noise and remove it using threshold-like techniques [27, 35, 68].

A recursive unsupervised segmentation approach applied on the decorrelated data space by the principal component analysis has been proposed by

Table 4.2 Overview of key document image enhancement techniques

Reference	Category	Short description	Remarks
Leung et al. [40]	Low contrast and uneven background illumination	Uses a generalized fuzzy operator preprocessed with histogram equalization and partially overlapped sub-block histogram equalization	Good contrast enhancement of extremely low contrast and low illuminated document images
Likforman et al. [43]	Low contrast and uneven background illumination	Combination of total variation regularization and nonlocal means filtering in order to enhance the quality of historical printed document images	It includes only one main parameter. Extensive tests based on OCR results
Nomura et al. [51]	Low contrast and uneven background illumination	Morphological operations are used to remove undesirable shapes called critical shadows on the background of document images	Does not need manual fine-tuning of parameters. Tested on degraded word images
Fadura et al. [23]	Bleed-through, shining, or shadow-through effects – non-registration	A recursive unsupervised segmentation approach applied on the decorrelated data space by the principal component analysis to remove bleed-through	Does not require any specific learning process or any input parameters
A. Tonazzini [65]	Bleed-through, shining, or shadow-through effects – non-registration	It is based on the projection of the RGB image into alternative color spaces that allow the enhancement of the main text areas	Does not require any intervention from the user side and no parameters need to be set
Dubois and Pathak [22]	Bleed-through, shining, or shadow-through effects – registration	An affine transformation is used to register the two sides; bleed-through is identified and replaced by the background color or intensity	Tested on documents generated under controlled conditions as well as on original manuscripts
Moghaddam and Cheriet [49]	Bleed-through, shining, or shadow-through effects – registration	It uses a new variational model with an extra term for reverse diffusion between the two sides of the document transferred to the wavelet domain	A blind (non-registration) version is also provided
Ajayi et al. [52]	Damaged characters or noisy background	An automated adaptive system to correct text degradations in typewritten documents based on lookup table classification algorithms	It is dependent on generating good quality ground truth images used for training
Drira et al. [21]	Damaged characters or noisy background	It suggests the application of the Partial Diffusion Equation (PDE) for enhancing text in degraded document images	This approach requires neither training nor segmentation steps
Shafait and Breuel [58]	Borders or parts of adjacent page – connected component analysis	It combines projection profile analysis with connected component removal to identify borders of noise regions	An efficient method simple to understand and implement

(continued)

Table 4.2 (continued)

Reference	Category	Short description	Remarks
Le et al. [38]	Borders or parts of adjacent page – projection profiles	It is based on classification of blank, textual, and non-textual rows and columns, location of border objects, and an analysis of projection profiles and crossing counts of textual squares	It uses several heuristics and it is based on several assumptions
Stamatopoulos et al. [62]	Borders or parts of adjacent page – projection profiles	It is based on projection profiles combined with a connected component labeling process. Signal cross-correlation is also used for verification	Experimental results on several camera document images, mainly historical
Haji et al. [29]	Borders or parts of adjacent page – projection profiles	A simultaneous document margin removal and skew correction based on corner detection in projection profiles	Applied to a collection of document images with different types of margin noise, layouts, and intensity levels
Avila and Lins [4]	Borders or parts of adjacent page – flood-fill	Invading and non-invading border detection algorithms. The non-invading border algorithm takes into account two parameters related to the nature of the documents in order to restrain flooding in the whole connected area	Tested on over 20,000 images and compared with several commercial tools (Scanfix, Leadtools, BlackIce, and Skyline Tools)
Avila and Lins [5]	Borders or parts of adjacent page – flood-fill	It is based on “flood-fill” component labeling and region adjacency graphs for removing noisy black borders in monochromatic images	Works even on images of torn-off documents or in the presence of irregular white stripes on the noise border frame
Fan et al. [24]	Borders or parts of adjacent page	It removes the black borders of scanned documents by reducing the resolution of the document image	It does not consider noisy text regions
Shafait et al. [59]	Borders or parts of adjacent page	It performs document image cleanup by detecting the page frame of the document	It can handle documents with a very large amount of noise with reasonable accuracy

Fadura et al. [23] for the enhancement of historical document images suffering from the bleed-through effect. This technique does not require any specific learning process or any input parameters. The stopping criterion for the proposed recursive approach has been determined empirically and set to a fixed number of iterations.

Recently, A. Tonazzini [65] has proposed some simple and fast procedures, based on the projection of the RGB image of a document into alternative color spaces. These color spaces, both fixed and self-adaptive, in many cases allow for the enhancement of the main text and the extraction of various features of the document. It is shown that even simpler arithmetic operations among the color channels can be effective for removing bleed-through, refocusing and improving the contrast of the foreground text.

Registration (Non-blind)-Based Methods

Registration-based methods require that the verso image of a page is available and precisely registered to the recto image. In order to register the verso with the recto image, several techniques have been proposed based on (i) intensity characteristics, intensity patterns are compared using correlation metrics; (ii) feature characteristics, correspondence between features such as points, lines, and contours is searched; (iii) transformation models; and (iv) spatial and frequency domain. In Dubois and Pathak's approach [22], an affine transformation is used to register the two sides, determining the parameters using an optimization method. Once the two sides have been registered, areas consisting primarily of bleed-through are identified and replaced by the background color or intensity. A new variational model has been introduced by Moghaddam and Cheriet [49] for the enhancement of degraded double-sided document images. This model has an extra term for reverse diffusion between the two sides of a document. By transferring the model to the wavelet domain and using the hard wavelet shrinkage, the solution of the model is obtained in a single step. The transformation to the wavelet domain is performed using the dual-tree complex wavelet transform which is a shift-invariant discrete wavelet transform.

Damaged Characters or Noisy Background

The enhancement of character areas and the elimination of noise in the background are usually a post-processing step of the binarization algorithms [27, 77]. Common operations that are applied in order to enhance the quality of a binary document image include masks, connected component analysis, morphological operations, as well as of shrink and swell operations.

Concerning the enhancement of grayscale or color images, Drira et al. [21] presented a study for restoring degraded text characters which consists of repairing the shapes of the features as well as extrapolating lost information. This study suggests the application of the Partial Diffusion Equation (PDE) for enhancing text in degraded document images. Ajayi et al. [52] introduced an automated adaptive system to correct text degradations in typewritten documents. It is based on lookup table classification algorithms and learns the corrections of patterns of text degradation in document images. Actual degraded historical documents are used for training a degradation model. The main limitation of this approach is that it is dependent on generating good quality ground truth images for the document collection to which it will be applied.

Borders or Parts of Adjacent Page

Border removal methods fall into the following categories:

Connected Component Analysis-Based Methods

The most common and easy to implement approach to eliminate marginal noise is to perform document cleaning by filtering out connected components based on their size and aspect ratio [53, 58]. However, when characters from the adjacent page are also present, they usually cannot be filtered out using only these features.

Projection Profile-Based Methods

Le et al. [38] proposed a method for border removal which is based on classification of blank, textual, and non-textual rows and columns, location of border objects, and an analysis of projection profiles and crossing counts of textual squares. Their approach uses several heuristics, and also it is based on the assumption that the page borders are very close to edges of images and borders are separated from image contents by a white space. However, this assumption is often violated.

Stamatopoulos et al. [62] have presented a technique for enhancing the document images captured by a digital camera by automatically detecting the document borders and cutting out noisy black borders as well as noisy text regions appearing from neighboring pages. This methodology is based on projection profiles combined with a connected component labeling process (see Fig. 4.12). Signal cross-correlation is also used in order to verify the detected noisy text areas.

A simultaneous document margin removal and skew correction based on corner detection in projection profiles has been proposed by Haji et al. [29]. The basic function of the algorithm is to find the corners which correspond to the page margins from the projection profiles of the input image. For a straight page, the leftmost and rightmost sharp corners in the horizontal profile of the image correspond to the left and right margins, and the leftmost and rightmost sharp corners in the vertical profile of the image correspond to the upper and lower margins. For skewed pages, we observe that horizontal and vertical projection profiles have an isosceles trapezoidal shape.

"Flood-Fill" Algorithms

Avila and Lins [4] proposed the invading and non-invading border detection algorithms which work as "flood-fill" algorithms. The invading algorithm assumes that the noisy black border does not invade the black areas of the document. It moves outside from the noisy surrounding borders towards the document. In the case that the document text region is merged to noisy black borders, the whole area, including the part of text region, is flooded and removed. Contrarily, the non-invading border algorithm assumes that noisy black border merges with document information. In order to restrain flooding in the whole connected area, it takes into account two parameters related to the nature of the documents. These are the maximum size of a segment belonging to a document and the maximum distance between lines. Also, Avila and Lins [5] proposed an algorithm based on "flood-fill" component labeling

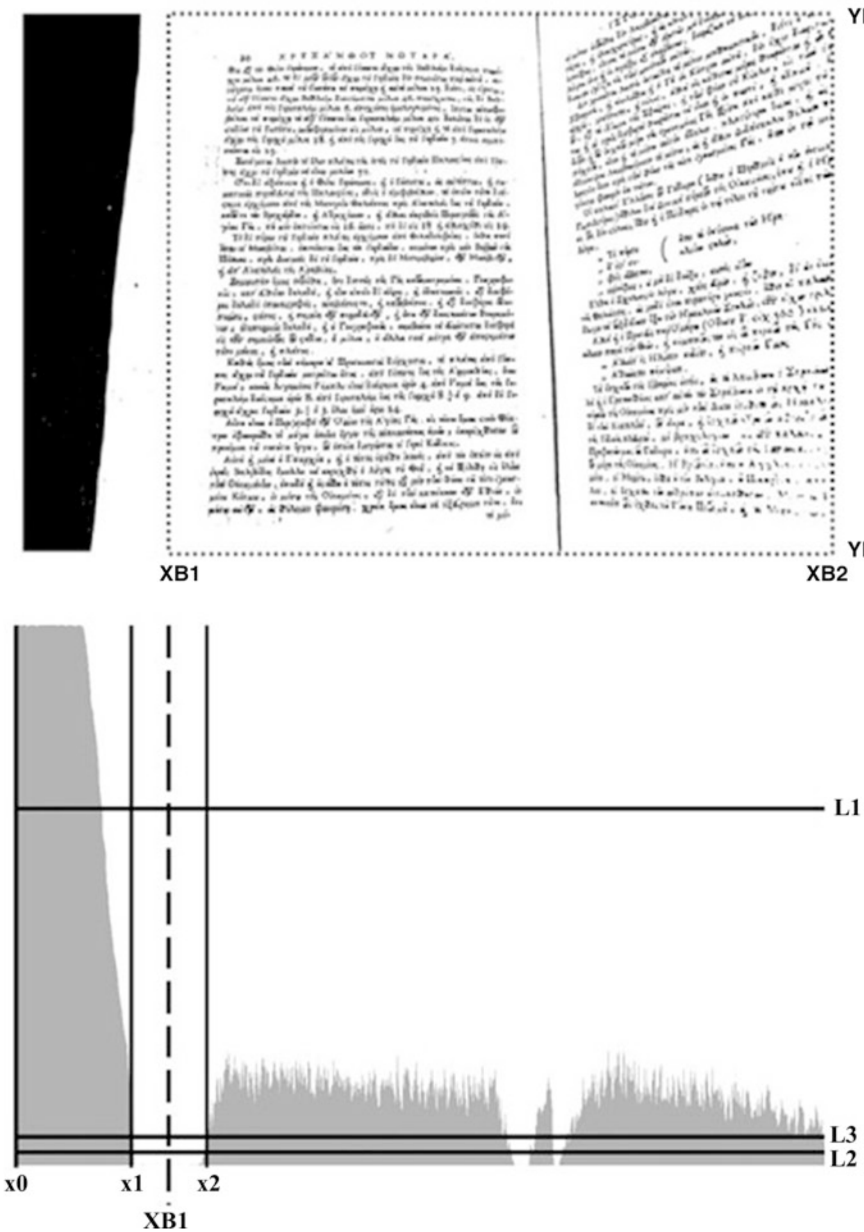


Fig. 4.12 Example of vertical projection profile used for border removal

and region adjacency graphs for removing noisy black borders in monochromatic images. The proposed algorithm encompassed five steps: flooding, segmentation, component labeling, region adjacency graph generation, and noise black border removal.

Other Methods

Fan et al. [24] proposed a method for removing noisy black regions overlapping the text region, but do not consider noisy text regions. They proposed a scheme to remove the black borders of scanned documents by reducing the resolution of the document image. This approach consists of two steps, marginal noise detection and marginal noise deletion. Marginal noise detection consists of three steps: (1) resolution reduction, (2) block splitting, and (3) block identification. Marginal noise detection makes the textual part of the document disappear leaving only blocks to be classified either as images or borders by a threshold filter. Marginal noise has to be deleted after it has been detected. The deletion process should be performed on the original image instead of the reduced image. The block classification is used to segment the original image, removing the noisy black borders.

Shafait et al. [59] proposed a new perspective for document image cleanup by detecting the page frame of the document. The goal of page frame detection is to find the actual page contents area, ignoring marginal noise along the page border (see Fig. 4.13). First, a geometric model is built for the page frame of a scanned structured document (journal article, book, magazine). Then, a geometric matching method is used to find the globally optimal page frame with respect to a defined quality function.

Document Image Normalization

Document digitization with either flatbed scanners or camera-based systems results in document images which often suffer from skew, warping, and perspective distortions that deteriorate the performance of current OCR approaches. On the other hand, document orientation is not known a priori, and as a result, the text may be rotated by 90° or even be upside down. To this end, a document image normalization step is imperative in order to restore text areas horizontally aligned without any distortions as well as in 0° angle. This step also includes slant correction for the case of handwritten documents where there is often a deviation of the near-vertical strokes from the vertical direction.

Page Orientation

The first step of document image normalization concerns the detection of text orientation. Usually, page orientation techniques involve the detection of portrait or landscape orientation, where the page has a skew of around 0° or 90°, correspondingly, as well as the detection of right side up or upside down orientation where the page has a skew of around 0° or 180°, correspondingly. An overview of the key page orientation estimation techniques is given in Table 4.3.

Portrait/landscape orientation detection is mainly accomplished by using projection histograms as well as by counting black-to-white transitions.

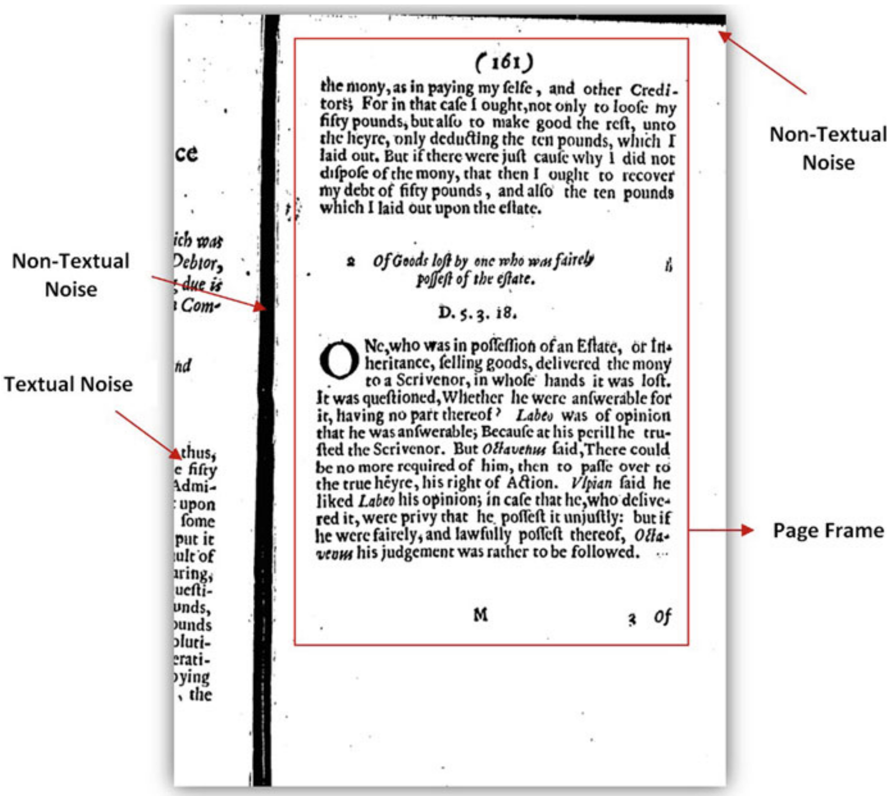


Fig. 4.13 An example of a page frame along with textual and non-textual noise

In [39], Le et al. use local analysis and projection histograms. The proposed methodology consists of three steps: textual and non-textual squares classification, textual squares page orientation estimation, and squares grouping. At the textual and non-textual squares classification step, the entire binary image is divided into squares based on a pyramidal image date structure, and each square is then classified as textual or non-textual based on several empirical conditions. At the textual squares page orientation step, the page orientation of each textual square is estimated by using a projection profiles or a square difference method. The projection profile method is based on an analysis of shapes of horizontal and vertical projection histograms, while the square difference method is based on the comparison between squared sums of horizontal and vertical projection histograms. According to the square difference method, text orientation is defined as follows:

$$
\begin{array}{lll}
S_H \geq aS_v & \Rightarrow \text{ text orientation } & = \text{portrait} \\
S_H < bS_v & \Rightarrow \text{ text orientation } & = \text{landscape} \\
bS_v \geq S_H > aS_v & \Rightarrow \text{ text orientation } & = \text{uncertain}
\end{array}
\tag{4.21}
$$

Table 4.3 Overview of key page orientation estimation techniques

Reference	Category	Short description	Remarks
Le et al. [39]	Portrait/ landscape orientation detection	Consists of three steps: textual and non-textual squares classification, textual squares page orientation estimation, and squares grouping. Orientation is estimated using a projection profile as well as a square difference method	Performs an accuracy rate of 99.9 % on a variety of more than 10,000 images
Yin [78]	Portrait/ landscape orientation detection	Counts the black-to-white transitions vertically and horizontally in the document image after applying a horizontal and vertical smoothing	Part of a skew detection and block classification process
Caprari [13]	Text page up/down orientation determination	Exploits an up/down asymmetry of passages of text composed of roman letters and Arabic numerals. Calculates the statistical excess of ascending over descending characters in the text areas	Tested on 226 pages (some of them containing noise) with 100 % success
Aradhye [3]	Text page up/down orientation determination	It analyzes the "open" portions of text blobs to determine the direction in which the open portions face	Adapted for roman and non-roman scripts, automated mail processing, checks, envelops, etc.
van Beusekom et al. [70]	One-step skew and orientation detection	A one-step skew and orientation detection method using a well-established geometric text line model	The effectiveness of the method is demonstrated using a publicly available dataset

where S_H and S_V the squared sums of horizontal and vertical projection histograms and a, b the upper and lower bound user-defined parameters. Each textual square is then assigned its mode weight as the total black pixels of a portrait or landscape mode textual square. At the squares grouping step, the page orientation of a binary image is determined using the notion of a pyramidal image data structure.

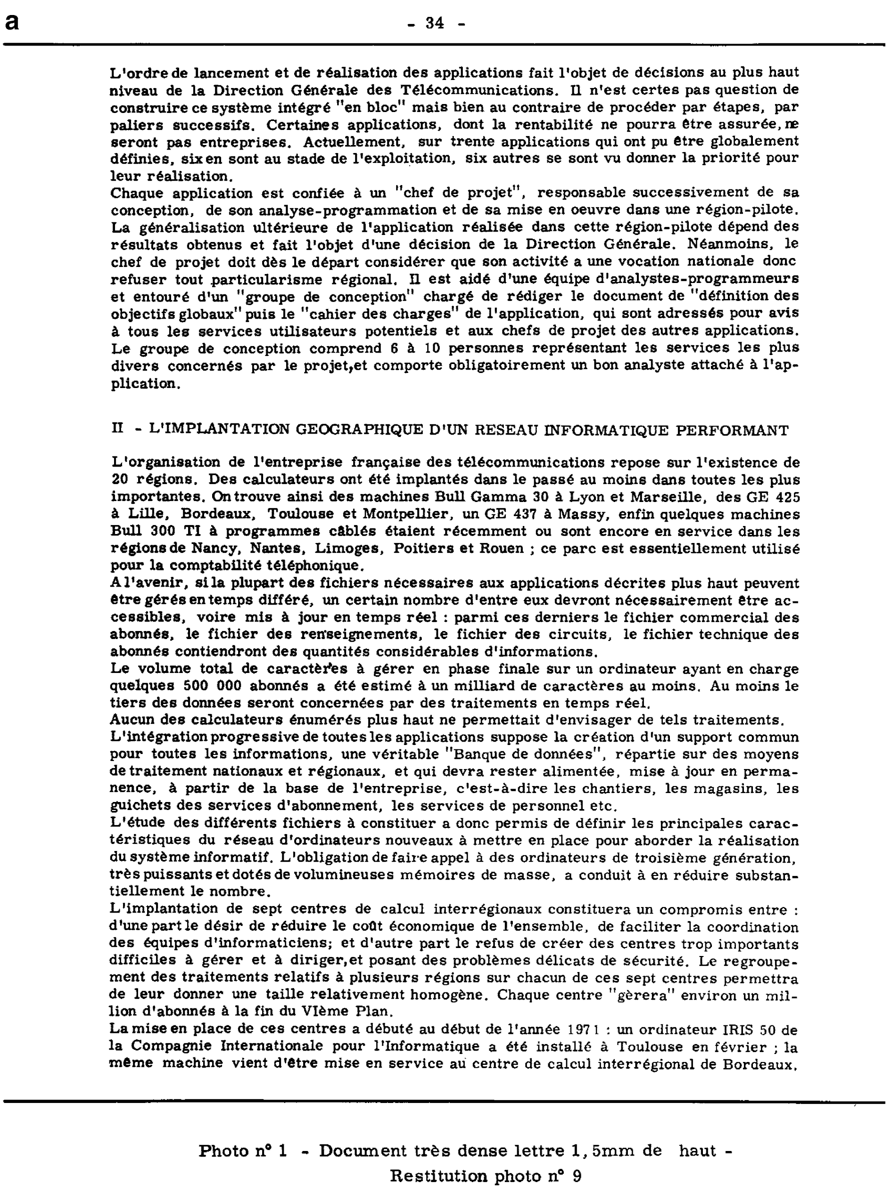
In method [78] proposed by Yin, the determination of text orientation is accomplished by counting the black-to-white transitions vertically and horizontally in the document image after applying a horizontal and vertical smoothing based on the run-length smoothing (RLSA) procedure [73]. As it is demonstrated in Fig. 4.14, we first apply a horizontal smoothing and calculate the black-to-white transitions in vertical direction (*TPix1*). Then, we first apply a vertical smoothing and calculate the black-to-white transitions in horizontal direction (*TPix2*). Text orientation is defined as follows:

$$\begin{aligned} TPix1 < TPix2 &\Rightarrow \text{ text orientation } = \text{portrait} \\ TPix1 \geq TPix2 &\Rightarrow \text{ text orientation } = \text{landscape} \end{aligned} \tag{4.22}$$

The orientation of the document image of Fig. 4.14 is detected as portrait since $TPix1 = 83{,}044 < TPix2 = 141{,}782$.

An algorithm for text page up/down orientation determination is presented in [13] by Caprari. The algorithm exploits an up/down asymmetry of passages of text

a



- 34 -

L'ordre de lancement et de réalisation des applications fait l'objet de décisions au plus haut niveau de la Direction Générale des Télécommunications. Il n'est certes pas question de construire ce système intégré "en bloc" mais bien au contraire de procéder par étapes, par paliers successifs. Certaines applications, dont la rentabilité ne pourra être assurée, ne seront pas entreprises. Actuellement, sur trente applications qui ont pu être globalement définies, six en sont au stade de l'exploitation, six autres se sont vu donner la priorité pour leur réalisation.
Chaque application est confiée à un "chef de projet", responsable successivement de sa conception, de son analyse-programmation et de sa mise en oeuvre dans une région-pilote. La généralisation ultérieure de l'application réalisée dans cette région-pilote dépend des résultats obtenus et fait l'objet d'une décision de la Direction Générale. Néanmoins, le chef de projet doit dès le départ considérer que son activité a une vocation nationale donc refuser tout particularisme régional. Il est aidé d'une équipe d'analystes-programmeurs et entouré d'un "groupe de conception" chargé de rédiger le document de "définition des objectifs globaux" puis le "cahier des charges" de l'application, qui sont adressés pour avis à tous les services utilisateurs potentiels et aux chefs de projet des autres applications. Le groupe de conception comprend 6 à 10 personnes représentant les services les plus divers concernés par le projet, et comporte obligatoirement un bon analyste attaché à l'application.

II - L'IMPLANTATION GEOGRAPHIQUE D'UN RESEAU INFORMATIQUE PERFORMANT

L'organisation de l'entreprise française des télécommunications repose sur l'existence de 20 régions. Des calculateurs ont été implantés dans le passé au moins dans toutes les plus importantes. On trouve ainsi des machines Bull Gamma 30 à Lyon et Marseille, des GE 425 à Lille, Bordeaux, Toulouse et Montpellier, un GE 437 à Massy, enfin quelques machines Bull 300 TI à programmes câblés étaient récemment ou sont encore en service dans les régions de Nancy, Nantes, Limoges, Poitiers et Rouen ; ce parc est essentiellement utilisé pour la comptabilité téléphonique.
A l'avenir, si la plupart des fichiers nécessaires aux applications décrites plus haut peuvent être gérés en temps différé, un certain nombre d'entre eux devront nécessairement être accessibles, voire mis à jour en temps réel : parmi ces derniers le fichier commercial des abonnés, le fichier des renseignements, le fichier des circuits, le fichier technique des abonnés contiendront des quantités considérables d'informations.
Le volume total de caractères à gérer en phase finale sur un ordinateur ayant en charge quelques 500 000 abonnés a été estimé à un milliard de caractères au moins. Au moins le tiers des données seront concernées par des traitements en temps réel.
Aucun des calculateurs énumérés plus haut ne permettait d'envisager de tels traitements.
L'intégration progressive de toutes les applications suppose la création d'un support commun pour toutes les informations, une véritable "Banque de données", répartie sur des moyens de traitement nationaux et régionaux, et qui devra rester alimentée, mise à jour en permanence, à partir de la base de l'entreprise, c'est-à-dire les chantiers, les magasins, les guichets des services d'abonnement, les services de personnel etc.
L'étude des différents fichiers à constituer a donc permis de définir les principales caractéristiques du réseau d'ordinateurs nouveaux à mettre en place pour aborder la réalisation du système informatif. L'obligation de faire appel à des ordinateurs de troisième génération, très puissants et dotés de volumineuses mémoires de masse, a conduit à en réduire substantiellement le nombre.
L'implantation de sept centres de calcul interrégionaux constituera un compromis entre : d'une part le désir de réduire le coût économique de l'ensemble, de faciliter la coordination des équipes d'informaticiens; et d'autre part le refus de créer des centres trop importants difficiles à gérer et à diriger, et posant des problèmes délicats de sécurité. Le regroupement des traitements relatifs à plusieurs régions sur chacun de ces sept centres permettra de leur donner une taille relativement homogène. Chaque centre "gèrera" environ un million d'abonnés à la fin du VIème Plan.
La mise en place de ces centres a débuté au début de l'année 1971 : un ordinateur IRIS 50 de la Compagnie Internationale pour l'Informatique a été installé à Toulouse en février ; la même machine vient d'être mise en service au centre de calcul interrégional de Bordeaux.

Photo n° 1 - Document très dense lettre 1,5mm de haut -
Restitution photo n° 9

Fig. 4.14 (continued)

composed of roman letters and Arabic numerals. Up/down text asymmetry is a result of the statistical excess of ascending characters over descending characters in the text areas. Specifically, for lower-case letters in common written English, the frequencies of occurrence of letters in the top, middle, and bottom rows are 26.5, 67.25, and 6.25 %, respectively. In order to calculate the document up/down

b

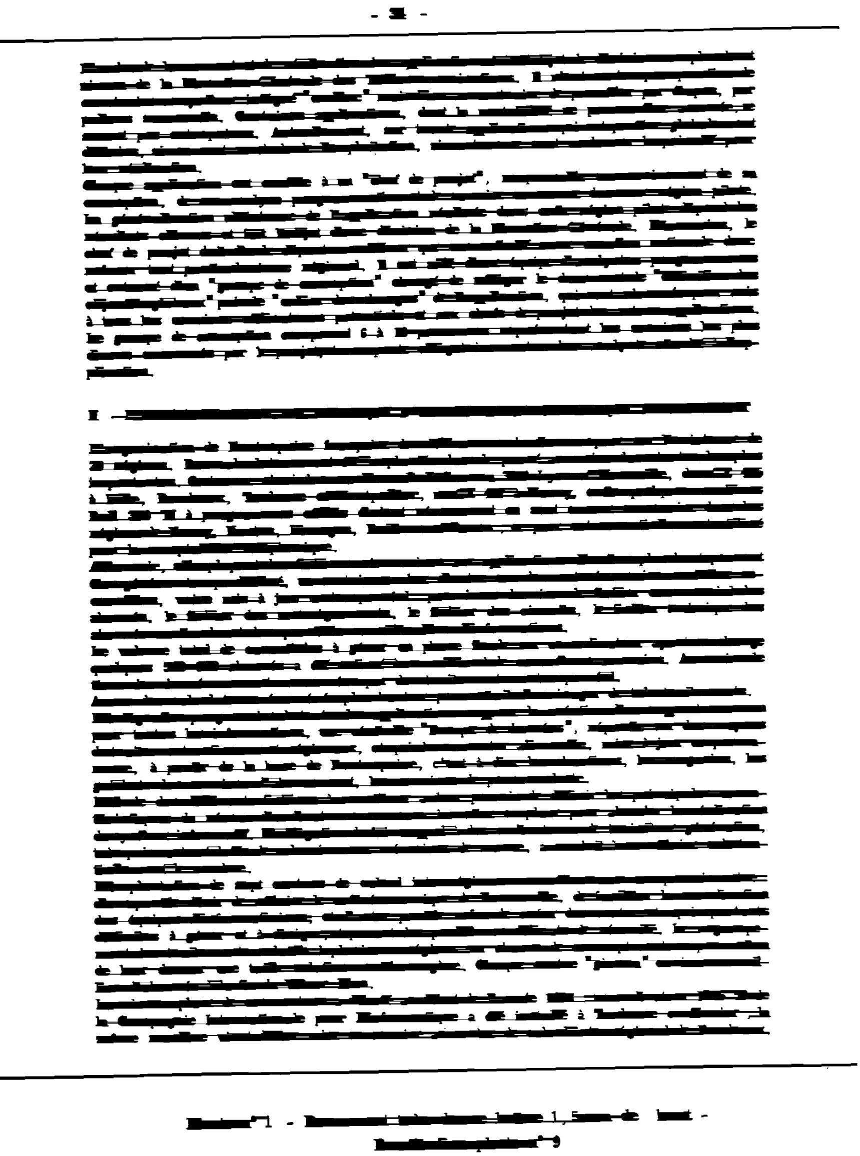


Fig. 4.14 (continued)

asymmetry, horizontal projections are used. A method for determining up/down orientation of text in roman and non-roman scripts is presented in [3] by Aradhye. The method analyzes the "open" portions of text blobs to determine the direction in which the open portions face. By determining the respective densities of blobs

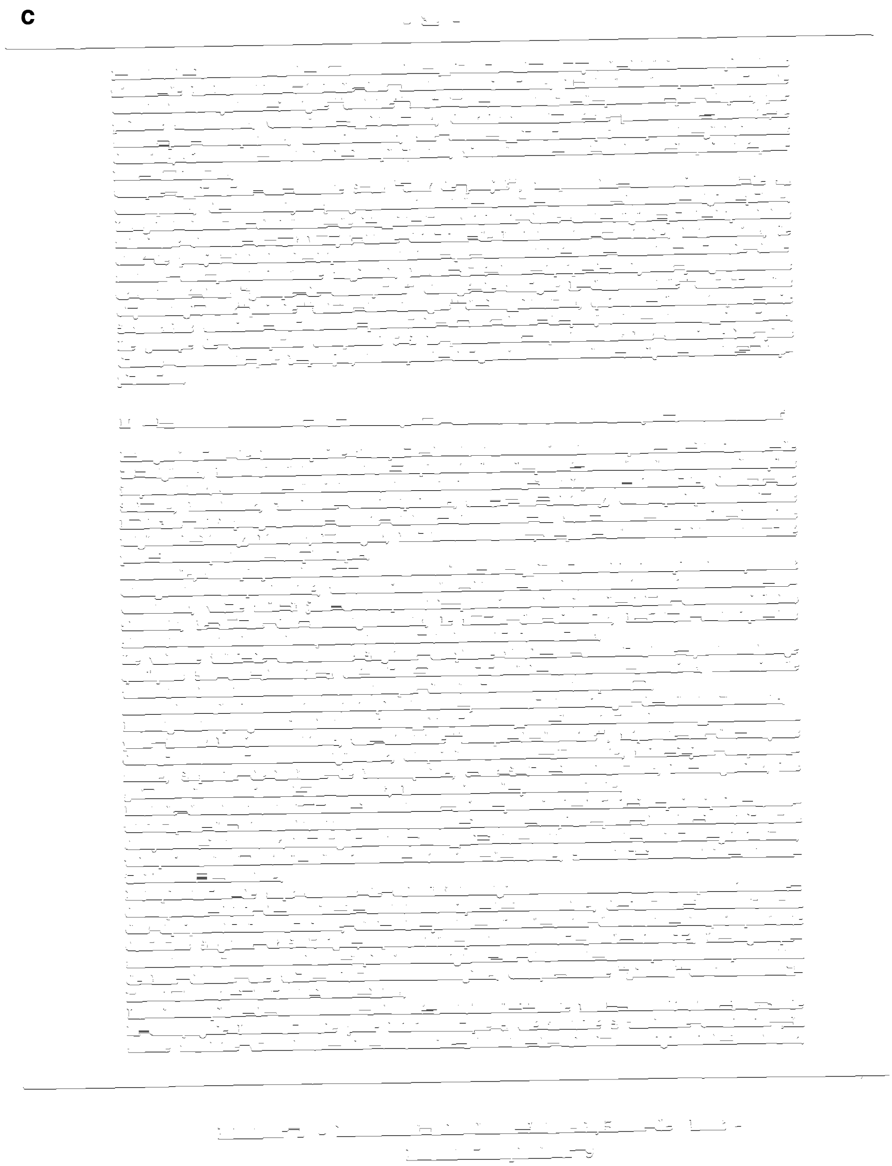


Fig. 4.14 (continued)

opening in a pair of opposite directions (e.g., right or left), the method can establish the direction in which the text as a whole is oriented. First, a method for determining the up/down orientation of roman text based on the asymmetry in the openness of most roman letters in the horizontal direction is presented. For non-roman text,

d

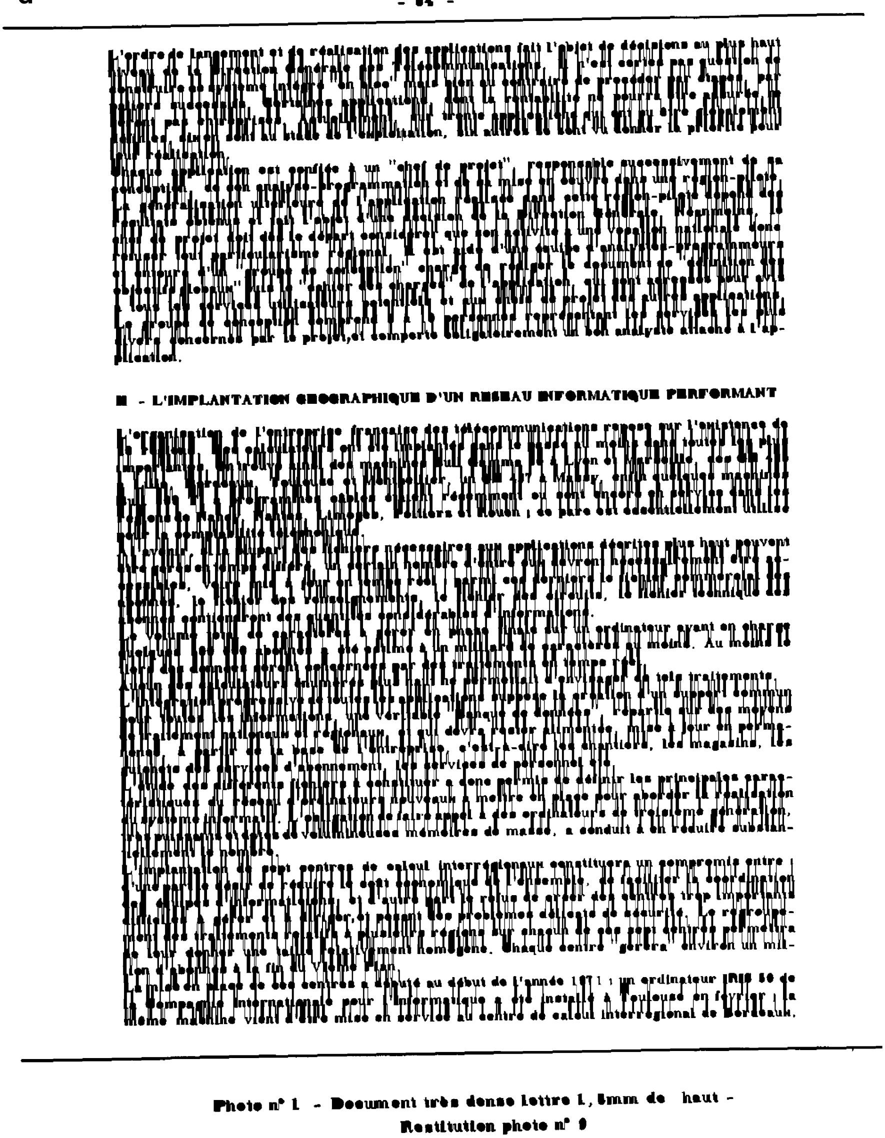


Fig. 4.14 (continued)

e

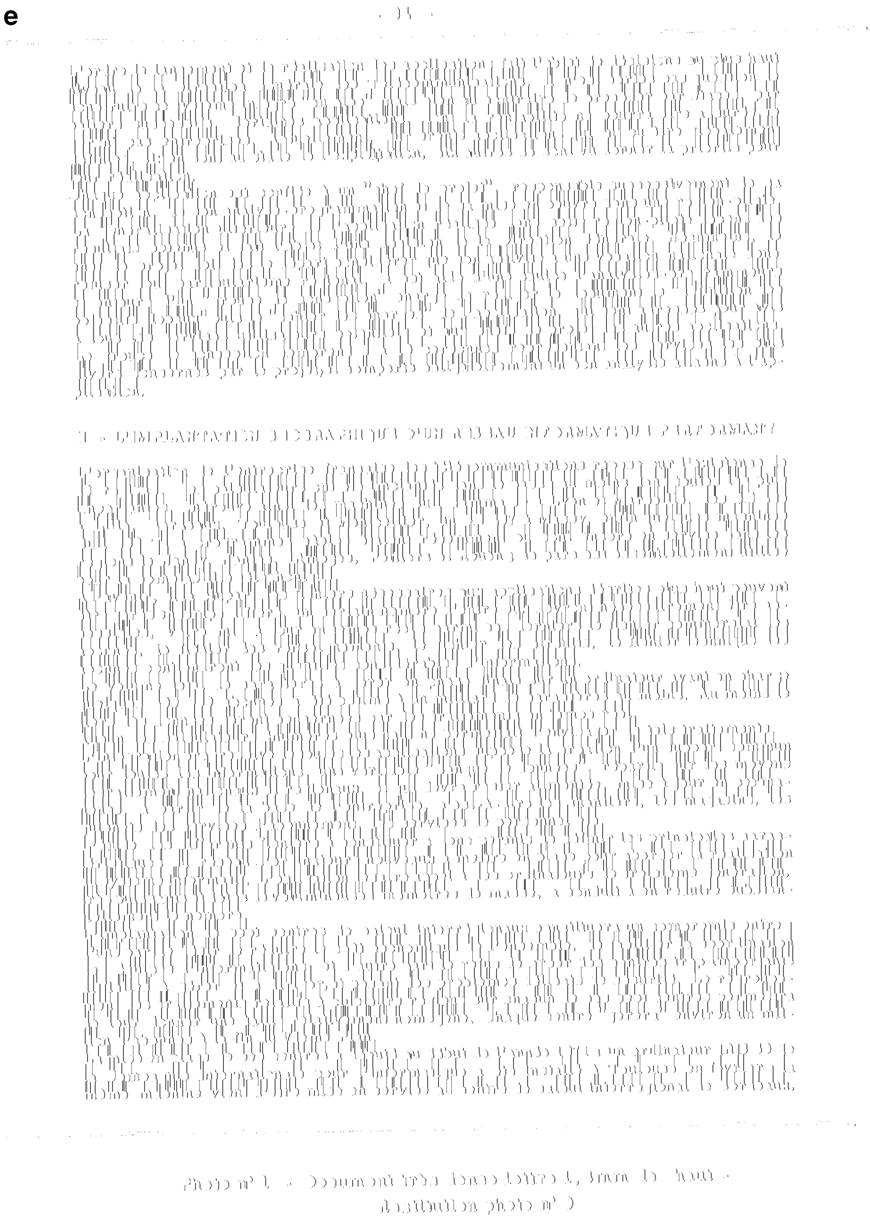


Fig. 4.14 Page orientation detection based on black-to-white transitions [78]. (**a**) Original binary image; (**b**) image after horizontal smoothing; (**c**) pixels that correspond to black-to-white transitions in vertical direction of the horizontally smoothed image (*TPix1* = total number of pixels = 83,044); (**d**) image after vertical smoothing; (**e**) pixels that correspond to black-to-white transitions in horizontal direction of the vertically smoothed image (*TPix2* = total number of pixels = 141,782)

a method that determines a direction that is the most asymmetric is provided. This direction is the most useful for the determination of text orientation, given a training dataset of documents of known orientation. This work can be adapted for use in automated mail processing or to determine the orientation of checks in automated teller machine envelopes, scanned or copied documents, documents sent via facsimile, etc.

A one-step skew and orientation detection method has been also proposed based on maximization of variance of transition counts (pixel from black to white or white to black) or on text line detection using a geometric text line model ([70] by van Beusekom et al.).

Deskew and Deslant

Document skew is often introduced during the document capturing process and may seriously affect the performance of subsequent stages of segmentation and recognition. An overview of the key document image skew detection techniques is given in Table 4.4.

Document skew detection methodologies proposed in the literature (see survey paper [31]) usually assume the skew angle range to be from $-5°$-to $5°$-and fall broadly into the following categories:

Projection Profile-Based Skew Detection Methods

According to these techniques, a series of horizontal projection profiles are calculated as we rotate the document page at a range of angles. The optimization of an objective function for a given skew angle leads to the actual document skew. As it can be observed in Fig. 4.15, the horizontal profiles tend to have consecutive local minima and maxima (discrete hills that correspond to text lines) as the document is rotated near $-\theta_d$ angle, where θ_d is the skew of the original document image. As an objective function that needs to be minimized, we can use the energy function $A(\theta)$ which is defined as follows:

$$A(\theta) = \sum_{i=1}^{m} c_i^2(\theta) \tag{4.23}$$

where m is the number of horizontal profile bins and θ the angle we rotate the document. At the example of Fig. 4.15, the original image has a skew of $\sim -3°$ and we calculate $A(0°) = 5.397.158$, $A(1°) = 5.537.790$, $A(2°) = 6.890.72653$, and $A(3°) = 8.927.232$.

In order to reduce high computational cost as well as to optimize search strategy, several variations of this method have been proposed. Baird [8] proposed a technique to minimize the number of points to be projected. For each connected component, only the midpoint of the bottom side of the bounding box is projected. The sum of the squares of the profile bins is used as the energy function to be

Table 4.4 Overview of key document image skew detection techniques

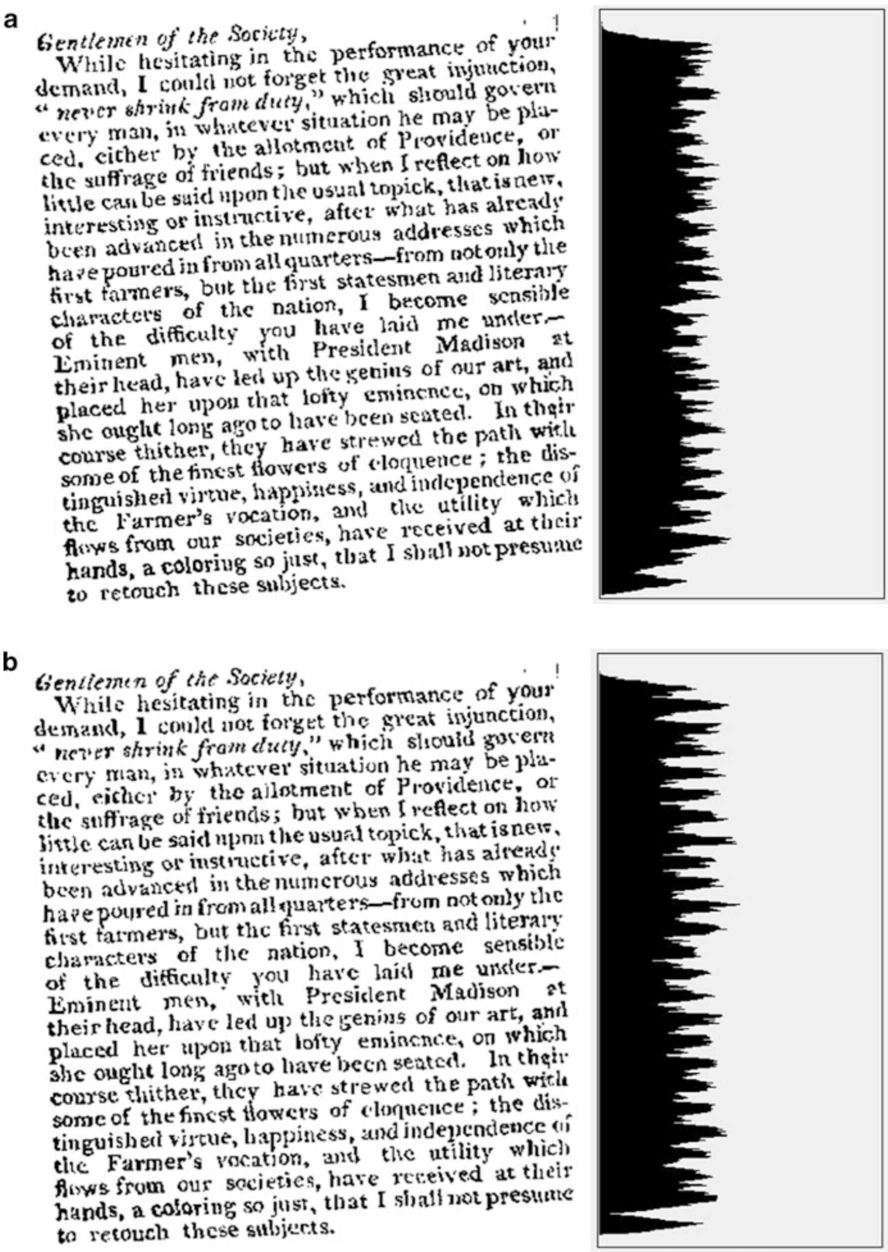
Reference	Category	Short description	Remarks
Baird [8]	Projection profiles	For each connected component, only the midpoint of the bottom side of the bounding box is projected. The sum of the squares of the profile bins is used as the energy function to be optimized	An iterative coarse-to-fine process was also introduced in order to speed up the process
Akiyama and Hagita [1]	Projection profiles	The image is partitioned vertically into a number of strips. The horizontal projection profile is then calculated for each strip. Correlation of the profiles of the neighboring strips	Fast but not so accurate results
Li et al. [41]	Projection profiles	Using wavelet decomposition, the horizontal structure of the document image is extracted and emphasized, which improves the accuracy of profile projection method	Fast, language independent, and can deal with document images with complex layouts and multiple font size
Amin and Fischer [2]	Hough transform	Connected components of similar dimensions are grouped together. Hough transform is performed after dividing each group of connected components into vertical segments	Fast and accurate for angles of up to 45°
Singh et al. [60]	Hough transform	Accelerates the application of Hough transform by reducing the number of image pixels using a modified form of block adjacency graph	Significant enhancement of the speed without affecting the skew detection accuracy
O'Gorman [53]	Nearest-neighboring clustering	Nearest-neighbor clustering uses the K-neighbors for each connected component	The calculated histogram peak may not be very accurate
Lu and Tan [45]	Nearest-neighboring clustering	Size restriction is introduced to the detection of nearest-neighbor pairs. The slopes of the chains with a largest possible number of nearest-neighbor pairs are calculated	Able to deal with documents of different scripts such as English, Tamil, and Chinese
Cao and Li [11]	Nearest-neighboring clustering	The bottom center of the bounding box of a connected component is regarded as an eigen-point. Skew is determined according to the relations between the successive eigen-points in every text line	It reduces the computing complexity and achieves better precision

(*continued*)

Table 4.4 (continued)

Reference	Category	Short description	Remarks
Yan [76]	Cross-correlation	The accumulated correlation for many pairs of lines is calculated. The optimal shift and the distance between each pair of lines are used for skew estimation	It can be used for binary, grayscale, and color images
Gatos et al. [26]	Cross-correlation	The image is first smoothed and then only the information existing in two or more vertical lines is used for skew detection	It can also be used for text line detection
Liu et al. [44]	Segmentation	It is based on detecting the borderlines of objects, either in text or non-text regions. Borderlines are extracted from the borders of large connected components by using a run-length-based method	Efficient for complex documents with horizontal and vertical text layout, predominant non-text regions, or sparse text regions
Chou et al. [17]	Segmentation	It is based on piecewise coverings of objects by parallelograms. The angle at which objects are best covered corresponds to the document skew angle	It is vulnerable to images that are rich in background noise
Fan et al. [25]	Segmentation	The Rectangular Active Contour Model (RAC Model) is used by assuming the boundary of a document image's content area as a rectangle represented by a five-tuple	Robust to noise and does not require any de-noising preprocessing
Chen and Wang [15]	Color documents	It determines the variation of color-transition count at each angle and the angle of maximal variation is regarded as the skew angle	Tested on 100 color images rotated from $-45°$ to $45°$
Makridis et al. [47]	Color documents	It consists of four main stages: color reduction, text localization, document binarization, and skew correction	It can be applied to color, grayscale, and binary documents

optimized. In order to speed up the search of the optimum angle, an iterative coarse-to-fine process was also introduced. In [1], Akiyama and Hagita proposed to partition the image vertically into a number of strips. The horizontal projection profile is then calculated for each strip. The skew angle is estimated from the correlation of the profiles of the neighboring strips. This yields fast but not so accurate result. A projection profile-based method using wavelet decompositions is presented in [41] by Li et al. Using the wavelet decomposition, the horizontal structure of the document image can be extracted and emphasized, which improves the accuracy of profile projection method. On the other hand, the computational cost is less than one tenth of that of the original after 2-level wavelet decomposition. Experimental results show that the proposed algorithm is language independent and can deal with document images with complex layouts and multiple font size.


a

Gentlemen of the Society,

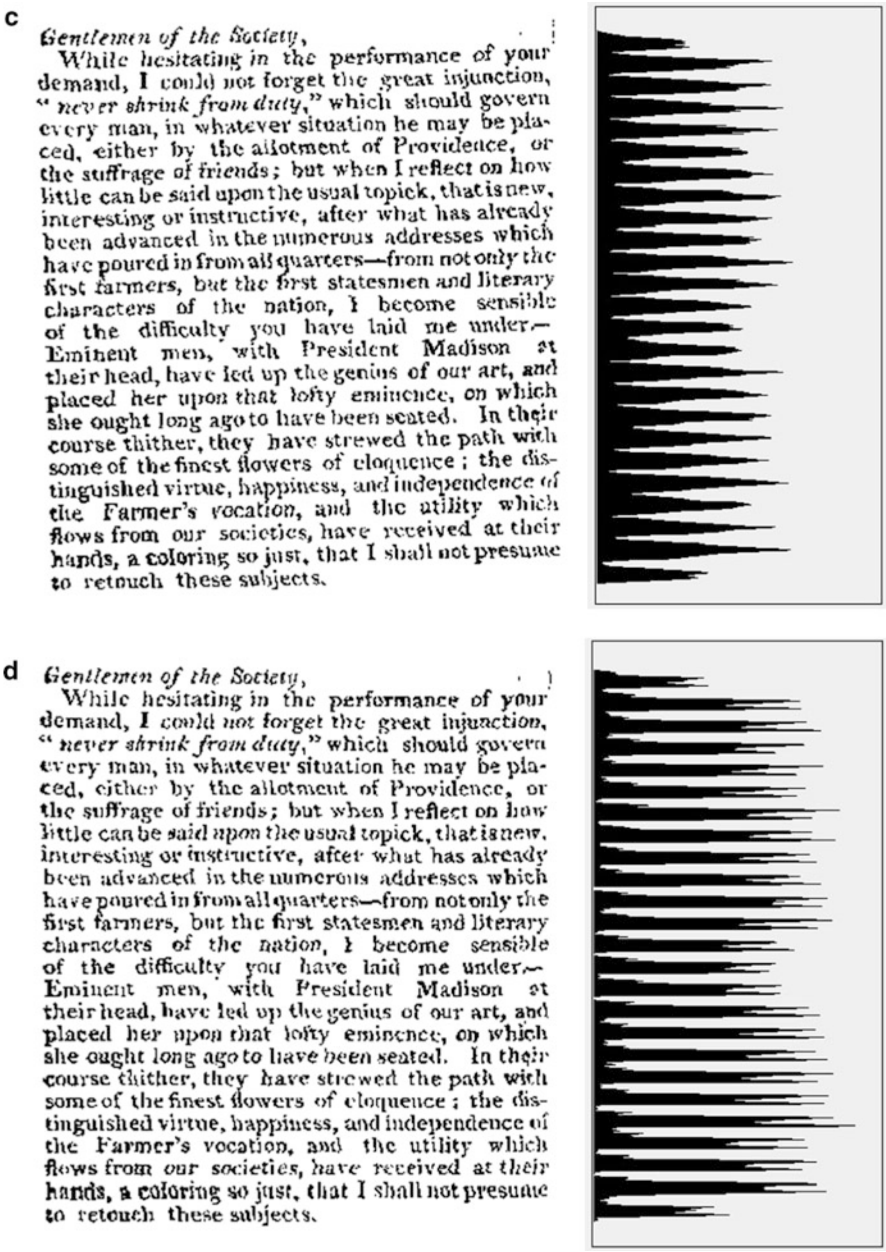
While hesitating in the performance of your demand, I could not forget the great injunction, " *never shrink from duty*," which should govern every man, in whatever situation he may be placed, either by the allotment of Providence, or the suffrage of friends; but when I reflect on how little can be said upon the usual topick, that is new, interesting or instructive, after what has already been advanced in the numerous addresses which have poured in from all quarters—from not only the first farmers, but the first statesmen and literary characters of the nation, I become sensible of the difficulty you have laid me under.—Eminent men, with President Madison at their head, have led up the genius of our art, and placed her upon that lofty eminence, on which she ought long ago to have been seated. In their course thither, they have strewed the path with some of the finest flowers of eloquence; the distinguished virtue, happiness, and independence of the Farmer's vocation, and the utility which flows from our societies, have received at their hands, a coloring so just, that I shall not presume to retouch these subjects.

b

Gentlemen of the Society,

While hesitating in the performance of your demand, I could not forget the great injunction, " *never shrink from duty*," which should govern every man, in whatever situation he may be placed, either by the allotment of Providence, or the suffrage of friends; but when I reflect on how little can be said upon the usual topick, that is new, interesting or instructive, after what has already been advanced in the numerous addresses which have poured in from all quarters—from not only the first farmers, but the first statesmen and literary characters of the nation, I become sensible of the difficulty you have laid me under.—Eminent men, with President Madison at their head, have led up the genius of our art, and placed her upon that lofty eminence, on which she ought long ago to have been seated. In their course thither, they have strewed the path with some of the finest flowers of eloquence; the distinguished virtue, happiness, and independence of the Farmer's vocation, and the utility which flows from our societies, have received at their hands, a coloring so just, that I shall not presume to retouch these subjects.

Fig. 4.15 (continued)



c

Gentlemen of the Society,

While hesitating in the performance of your demand, I could not forget the great injunction, " *never shrink from duty*," which should govern every man, in whatever situation he may be placed, either by the allotment of Providence, or the suffrage of friends; but when I reflect on how little can be said upon the usual topick, that is new, interesting or instructive, after what has already been advanced in the numerous addresses which have poured in from all quarters—from not only the first farmers, but the first statesmen and literary characters of the nation, I become sensible of the difficulty you have laid me under.—Eminent men, with President Madison at their head, have led up the genius of our art, and placed her upon that lofty eminence, on which she ought long ago to have been seated. In their course thither, they have strewed the path with some of the finest flowers of eloquence; the distinguished virtue, happiness, and independence of the Farmer's vocation, and the utility which flows from our societies, have received at their hands, a coloring so just, that I shall not presume to retouch these subjects.

d

Gentlemen of the Society,

While hesitating in the performance of your demand, I could not forget the great injunction, " *never shrink from duty*," which should govern every man, in whatever situation he may be placed, either by the allotment of Providence, or the suffrage of friends; but when I reflect on how little can be said upon the usual topick, that is new, interesting or instructive, after what has already been advanced in the numerous addresses which have poured in from all quarters—from not only the first farmers, but the first statesmen and literary characters of the nation, I become sensible of the difficulty you have laid me under.—Eminent men, with President Madison at their head, have led up the genius of our art, and placed her upon that lofty eminence, on which she ought long ago to have been seated. In their course thither, they have strewed the path with some of the finest flowers of eloquence; the distinguished virtue, happiness, and independence of the Farmer's vocation, and the utility which flows from our societies, have received at their hands, a coloring so just, that I shall not presume to retouch these subjects.

Fig. 4.15 Horizontal projection profiles of a document image rotated at angle θ: (**a**) original image having a skew of $\sim -3^{\circ}$ and rotated at 0°, (**b**) original image rotated at 1°, (**c**) original image rotated at 2°, and (**d**) original image rotated at 3°

Hough Transform-Based Skew Detection Methods

Since usually the document skew is also the skew of the document text lines, the use of the Hough transform (HT) for detecting straight lines has been proposed as a skew detection tool (see survey paper [31]). Each black pixel is mapped to the Hough space (ρ, θ) and the skew is estimated as the angle in the parameter space that gives the maximum sum of squares of the gradient along the ρ component. Several research efforts focus on improving the computational efficiency of this method. The image can be downsampled and transformed into a "burst image." The "burst image" is build by replacing each vertical black run with its length and placing this value in the bottommost pixel of the run. HT is then applied only to the pixels of the burst image that have value less than a threshold in order to discard non-text components. HT can be performed on a selected square of the document where only bottom pixels of candidate objects are preserved. According to the skew detection methodology proposed in [2] by Amin and Fischer, connected components of similar dimensions are grouped together and a skew angle is calculated for each group. HT is performed after dividing each group of connected components into vertical segments of approximately the width of one connected component and stores only the bottom rectangle in each segment. The skew angle for the entire page is then determined by choosing the most frequent angle, after applying a weighting factor to each angle to account for the number of points used in its calculation. In [60], the two major issues of the Hough transform, slow speed and high memory requirement, are addressed by Singh et al. The overall speed of skew detection is enhanced by using a preprocessing stage where the number of image pixels is greatly reduced using a modified form of block adjacency graph.

Nearest-Neighbor Clustering-Based Skew Detection Methods

According to these approaches, spatial relationships and mutual distances of connected components are used to estimate the page skew. The direction vector of all nearest-neighbor pairs of connected components is accumulated in a histogram and the peak in the histogram gives the dominant skew. This method is generalized in [53] by O'Gorman where nearest-neighbor clustering uses the K-neighbors for each connected component. Since connections between neighboring components may be made both within and across text lines, the calculated histogram peak may not be very accurate. An improved nearest-neighbor-based approach for skew estimation is presented in [45] by Lu and Tan. According to this approach, size restriction is introduced to the detection of nearest-neighbor pairs. Then, the chains with a largest possible number of nearest-neighbor pairs are selected, and their slopes are computed to give the skew angle of document image. In [11] by Cao and Li, the bottom center of the bounding box of a connected component is regarded as an eigen-point. According to the relations between the neighboring eigen-points in every text line, the eigen-points laid on the baseline are extracted as sample points. Then, these samples are adopted by the least squares method to calculate the baseline direction. The average of these baseline directions corresponds to the skew angle of the whole document image. To reduce the computing cost, a suitable subregion that only contains the pure text content is selected for the skew detection method.

Cross-Correlation-Based Skew Detection Methods

These approaches are based on measuring vertical deviations among foreground pixels along the document image in order to detect the page skew. Cross-correlation between lines at a fixed distance is used in [76] by Yan. This method is developed based on the following observation. The correlation between two vertical lines of a skewed document image is maximized if one line is shifted relatively to the other such that the character baseline levels for the two lines are coincident. The accumulated correlation for many pairs of lines can provide a precise estimation of the page skew. The skew angle can be determined from the optimal shift and the distance between each pair of lines. Cross-correlation $R_1(x, s)$ can be defined as

$$R_1(x, s) = \sum_{y=S}^{I_y - S} B(x + d, y + s) B(x, y) \tag{4.24}$$

where $B(x, y)$ the binary image with $B(x, y) = 1$ represents a dark pixel in a character and $B(x, y) = 0$ a bright pixel in the background, $x \in [0, I_x]$, $y \in [0, I_y]$, s the sift between two vertical lines of the image at x and $x + d$, and $s \in [-S, S]$. The accumulated cross-correlated function $R(s)$ is defined as follows:

$$R(s) = \sum_{x=0}^{I_x - d} R_1(x, s) \tag{4.25}$$

In [26] by Gatos et al., the image is first smoothed by a horizontal run-length algorithm [73], and then, only the information existing in two or more vertical lines is used for skew detection. Based only on this information, 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.

Segmentation-Based Skew Detection Methods

These methodologies first proceed to document image segmentation in order to detect image objects (e.g., text or non-text regions, text lines) and then estimate the skew angle based on analyzing the skew characteristics of each object. A skew detection method for complex document images based on detecting the borderlines of objects, either in text regions or in non-text regions, is proposed in [44] by Liu et al. Borderlines are extracted from the borders of large connected components in a document image by using a run-length-based method [73]. After filtering out nonlinear borderlines, a fast iteration algorithm is applied to optimize each linear borderline's directional angle. Finally, the weighted median value of all the directional angles is calculated as the skew angle of the whole document. A skew estimation algorithm based on piecewise coverings of objects by parallelograms was proposed in [17] by Chou et al. In this approach, the document image is divided into several non-overlapping regions and the objects within each region are covered by parallelograms at various angles. The angle at which objects are best covered corresponds to skew angle of the document. The Rectangular Active Contour Model (RAC Model) has been recently proposed in [25] by Fan et al. for skew detection by assuming the boundary of a document image's content area as a rectangle represented by a five-tuple. This assumption is satisfied with many document images, such as journals, magazines, and newspaper. In the model, a rectangular shape constraint is imposed on the zero-level set of a scalar Lipschitz continuous function in Chan–Vese Model (C–V Model). The variational method is applied to minimize the energy functional to get the parameters of the rectangle close to the boundary of the content region. The proposed algorithm estimates the skew angle with a global shape feature of content on a document image, so it is robust to noise and does not require any de-noising preprocessing.

A one-step skew and orientation detection method is proposed in [70] by van Beusekom et al. and is based on text line detection using a geometric text line model.

Skew Detection Methods for Color Documents

The skew detection and correction approach of [15] by Chen and Wang has been designed for color documents. This approach first determines variation of color-transition count at each angle and the angle of maximal variation is regarded as the skew angle. Then, a scanning-line model reconstructs the image. A technique for global and local skew detection in complex color documents is proposed in [47] by Makridis et al. It consists of four main stages: color reduction, text localization, document binarization, and skew correction. Skew correction is achieved by detecting the direction of connection of the connected components in the binary images.

By the term "character slant," we imply the angle in degrees clockwise from vertical at which the characters are drawn (see Fig. 4.16). Character slant estimation can be very helpful in handwritten text processing. Knowing the slant's value, we can correct it in order to facilitate processing and recognition. In addition to that, the character slant is thought to be very important information which can help to point out the writer of a text. Document slant may be uniform (see Fig. 4.16a) or nonuniform (see Fig. 4.16b) along the text line or even within the same word. An overview of key slant estimation techniques is given in Table 4.5.

Slant estimation methodologies proposed in the literature fall broadly into the following main categories.

Slant Estimation Methods that Analyze Near-Vertical Strokes

According to the approach of Bozinovich and Srihari [9], for a given word, all horizontal lines which contain at least one run of length greater than a parameter M (depending on the width of the letters) are removed. Additionally, all horizontal strips of small height are removed. By deleting these horizontal lines, only orthogonal window parts of the picture remain in the text. For each letter, the parts that remain (after the deletion) are those that contribute to the word slant. For each of these parts, we estimate the angle between the line indicated by the centers of gravity for its upper and lower halves and the vertical side of the page.

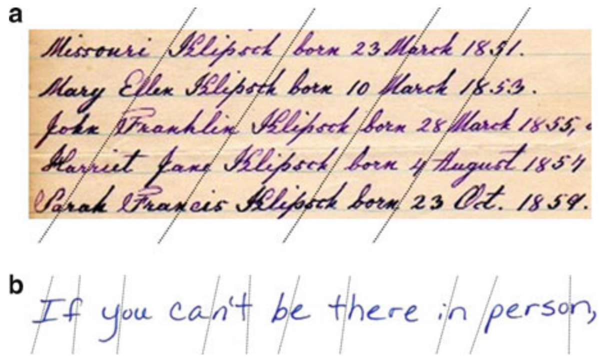


Fig. 4.16 Slanted handwriting examples: (**a**) uniform slanted text and (**b**) nonuniform slanted text

The mean value of these slants is the overall text's character slant. In the approach of Kim and Govindaraju [34], vertical and near-vertical lines are extracted by tracing chain code components using a pair of one-dimensional filter. Coordinates of the start and end points of each vertical line extracted provide the slant angle. Global slant angle is the average of all the angle of the lines, weighted by their length direction since the longer line gives more accurate angle than the shorter one.

Projection Profile-Based Slant Estimation Methods

Vinciarelli and Luettin [71] have proposed a deslanting technique based on the hypothesis that the word has no slant when the number of columns containing a continuous stroke is maximum. The word image is artificially slanted at different slant angles, and for each angle α, the following vertical profile is calculated:

$$H_\alpha(m) = \frac{h_\alpha(m)}{\Delta y_a(m)} \tag{4.26}$$

where $h_\alpha(m)$ is the vertical density (number of foreground pixels per column) in column m, and $\Delta y_\alpha(m)$ the distance between the highest and lowest pixel in the same column. If the column m contains a continuous stroke, $H_\alpha(m) = 1$, otherwise $H_\alpha(m) \in [0,1)$. For each angle α, the following function is calculated:

$$S(\mathrm{a}) = \sum_{\{i:H_\alpha(i)=1\}} h_\alpha(i)^2 \tag{4.27}$$

The angle α giving the highest value of S is taken as the slant estimate.

Kavallieratou et al. [33] have proposed a slant removal algorithm based on the use of the vertical projection profile of word images and the Wigner–Ville distribution (WVD). The word image is artificially slanted, and for each of the extracted word images, the vertical histogram as well as the WVD of these histograms is calculated. The curve of maximum intensity of the WVDs corresponds to the histogram with the most intense alternations and as a result to the dominant word slant.

Table 4.5 Overview of key slant estimation techniques

Reference	Category	Short description	Remarks
Bozinovich and Srihari [9]	Analysis of near-vertical strokes	After removing character parts that do not contribute to the word slant, the angle calculated in several image parts is used to estimate the global word slant	Several parameters have to be manually tuned
Kim and Govindaraju [34]	Analysis of near-vertical strokes	Vertical and near-vertical lines are extracted by tracing chain code components using a pair of one-dimensional filter. Global word angle is calculated by the weighted average of the angle of these lines	Part of a chain code-based handwritten word recognition system
Vinciarelli and Luettin [71]	Projection profiles	It is based on the hypothesis that the word has no slant when the number of columns containing a continuous stroke is maximum	It is shown to improve the recognition rate by 10 % relative to other normalization methods
Kavallieratou et al. [33]	Projection profiles	It is based on the use of the vertical projection profile of word images and the Wigner–Ville distribution (WVD) of this profile	Tested in English and Modern Greek samples of more than 500 writers
Kimura et al. [36]	Chain code	The average slant of the handwritten word is estimated using the chain code histogram of border pixels	The slant tends to be underestimated when the absolute of the slant is close or greater than 45°
Ding et al. [19]	Chain code	A modification of [36] using an 8-directional chain code	12- and 16-directional chain code are also examined but it is shown that the slant is overestimated
Ding et al. [20]	Chain code	Cumulative frequency distribution of chain code is used for local slant estimation	Several variations are also proposed in order to improve the accuracy

Chain Code-Based Slant Estimation Methods

Average slant of a handwritten word can be easily estimated using the chain code histogram of border pixels. According to Kimura et al. [36], the slant θ of a handwritten word can be estimated as follows:

$$\theta = \tan^{-1}\left(\frac{n_1 - n_3}{n_1 + n_2 + n_3}\right) \tag{4.28}$$

where n_i is the number of chain elements at an angle of $i \times 45°$ ("/" or "|" or "\"). Although this method usually gives good estimate of the word slant, the slant tends to be underestimated when the absolute of the slant is close or greater than 45°.

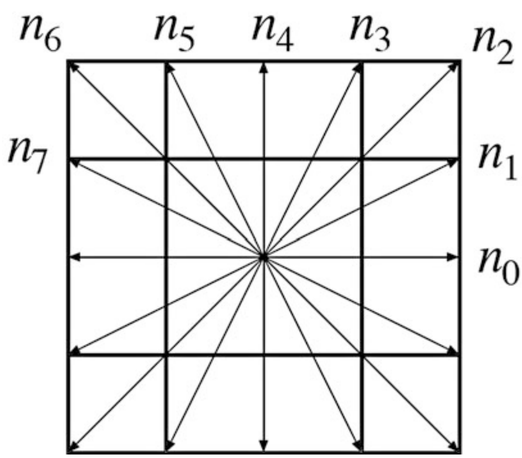


Fig. 4.17 8-directional quantization of chain code

To solve the problem, Ding et al. [19] proposed a slant detection method using an 8-directional chain code. Obtained chain code is then quantized to eight directions as shown in Fig. 4.17. Each direction is numbered 0–7 in counterclockwise, and n_i denotes the number of chain code elements in direction i (see Fig. 4.17). The slant estimator of the 8-directional chain code method is given as follows:

$$\theta = \tan^{-1}\left\{\frac{(2n_1 + 2n_2 + n_3) - (n_5 + 2n_6 + 2n_7)}{(n_1 + 2n_2 + 2n_3) + 2n_4 + (2n_5 + 2n_6 + n_7)}\right\} \tag{4.29}$$

where $(2n_1 + 2n_2 + n_3)$ is the sum of horizontal projection of the element 1, 2, 3; $(n_5 + 2n_6 + 2n_7)$ is the sum of horizontal projection of the element 5, 6, 7; and the denominator is the sum of vertical projection of the element 1–7. Chain code methods of 12 and 16 directions are also examined in [19], but the experimental results show that these methods tend to overestimate the slant.

A local slant estimation using cumulative frequency distribution of chain code has been proposed by Ding et al. [20]. Local slant estimator is defined as a function of horizontal coordinate x as follows:

$$\theta(x) = \tan^{-1}\left[\frac{n_1(x) - n_3(x)}{n_1(x) + n_2(x) + n_3(x)}\right] \tag{4.30}$$

where $n_i(x)$ $(i = 1, 2, 3)$ is the frequency distribution of chain code elements at direction of $i \times 45°$ in $[x - \delta_x, x + \delta_x]$ calculated as follows:

$$n_i(x) = s_i(x + \delta_x) - s_i(x - \delta_x - 1) \tag{4.31}$$

where $s_i(x)$ is the cumulative frequency distribution number of chain code elements at direction of $i \times 45°$ in $[0, x]$. Parameter δ_x is determined experimentally depending on the input image. Then, $\tan\theta(x)$ is smoothed by the mean filter of adjacent three pixels. To improve the accuracy of local slant estimation, three variations of the above idea (simple iterative method, high speed iterative method, and 8-directional chain code method) are also proposed in [20].

Dewarping

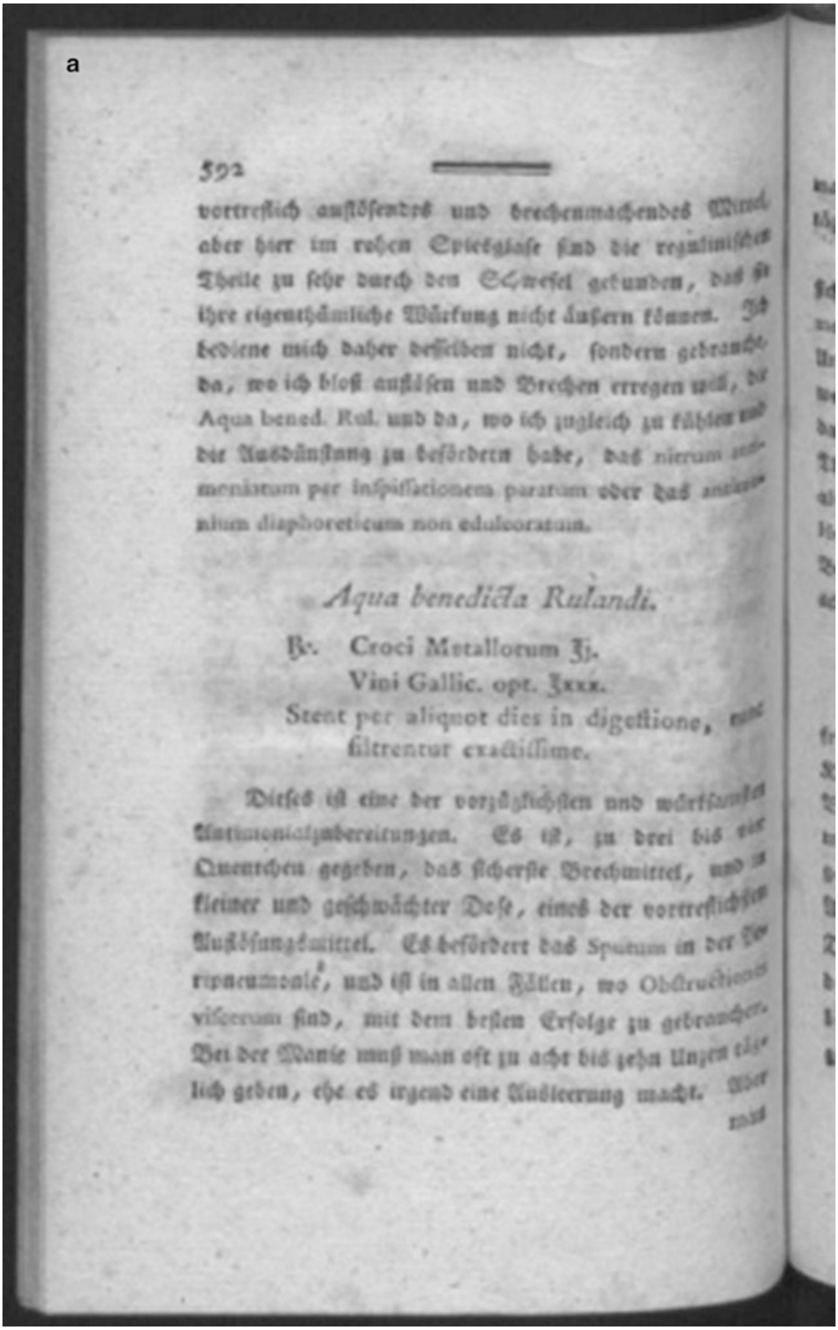
Document image acquisition by a flatbed scanner or a digital camera often results in several unavoidable image distortions (see Fig. 4.18) due to the form of printed material (e.g., bounded volumes), the camera setup, or environmental conditions (e.g., humidity that causes page shrinking). Text distortions not only reduce document readability but also affect the performance of subsequent processing such as document layout analysis and character recognition. Many different techniques have been proposed for document image rectification that can be classified into two main categories based on (1) 3-D document shape reconstruction [12,42,80] and (2) 2-D document image processing [10,37,46,48,63,75,79]. Techniques of the former category obtain the 3-D information of the document image using special setup or reconstruct the 3-D model from information existing in document images. On the other hand, techniques in the latter category do not depend on auxiliary hardware or prior information but they only rely on 2-D information. An overview of document image dewarping techniques is given in Table 4.6.

3-D Document Shape-Based Methods

In this category, rectification techniques rely upon extraction of the 3-D information of the document and they can be further divided into two subcategories. Techniques of the first subcategory obtain the 3-D shape of the document image using special equipment such as laser scanners, stereo cameras, or structured light setups. The dependence on special equipment prevents these techniques from being used in an unconstrained environment. On the other hand, techniques of the second subcategory reconstruct the 3-D model from information existing in document images. Cao et al. [12] propose a method to rectify warping distortions in document images by constructing a cylinder model. Apart from the cylinder shape assumption, they also have a limitation on the pose that requires the image plane to be parallel to the generatrix of the page cylinder. Liang et al. [42] model the page surface by curved developable surfaces to estimate the 3-D shape of the page using texture flow fields. This method is based on the assumptions that the document is either flat or smoothly curved and the camera is a standard pinhole camera. Finally, some methods are based on a shape-from-shading formulation in order to reconstruct the 3-D shape of the document's surface (e.g., L. Zhang et al. [80]). These techniques require knowledge of lighting which in most of the cases is unknown.

2-D Document Image Processing-Based Methods

In this category, rectification techniques rely on the use of 2-D information available in document images. The majority of these rectification techniques are based on the detection of distorted text lines at the original document image which is a well-known hard task. Some of these techniques propose a method to straighten distorted text lines by fitting a model to each text line. Lavialle et al. [37] use an analytical


a

592

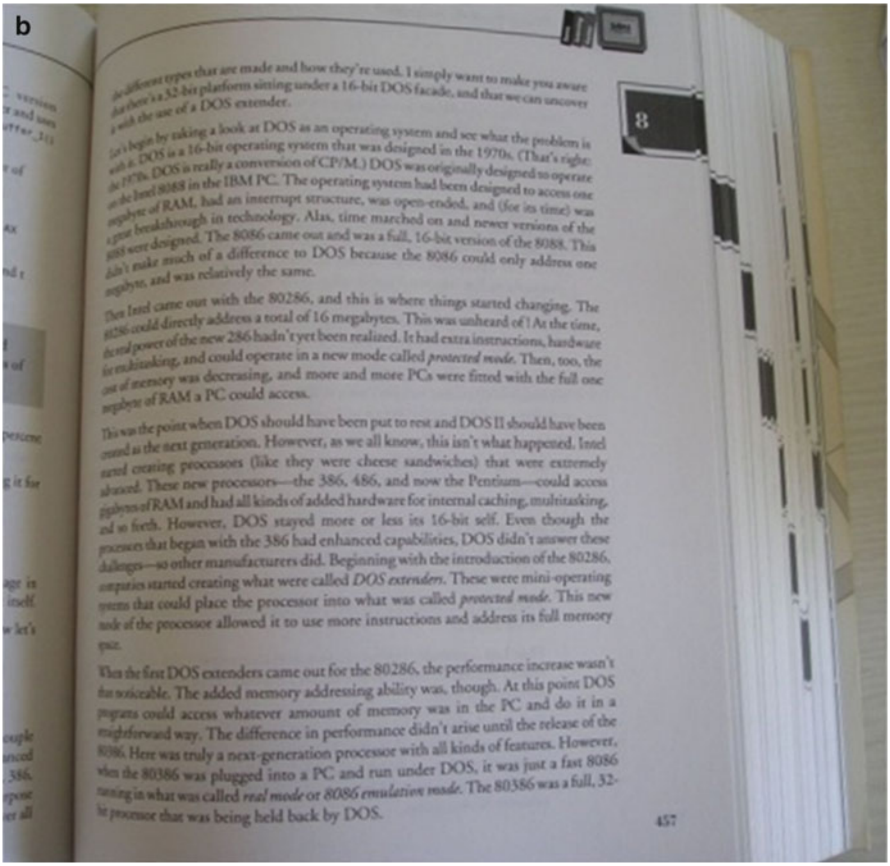
vortreflich auflösendes und brechenmachendes Mittel, aber hier im rohen Spiesglase sind die regulinischen Theile zu sehr durch den Schwefel gebunden, daß sie ihre eigenthümliche Würkung nicht äußern können. Ich bediene mich daher desselben nicht, sondern gebrauche da, wo ich bloß auflösen und Brechen erregen will, die Aqua bened. Rul. und da, wo ich zugleich zu kühlen und die Ausdünstung zu befördern habe, das nitrum antimoniatum per inspissationem paratum oder das antimonium diaphoreticum non edulcoratum.

Aqua benedicta Rulandi.

℞. Croci Metallorum ℥j.
Vini Gallic. opt. ℥xxx.
Stent per aliquot dies in digestione, tunc filtrentur exactissime.

Dieses ist eine der vorzüglichsten und würksamsten Antimonialzubereitungen. Es ist, zu drei bis vier Quentchen gegeben, das sicherste Brechmittel, und in kleiner und geschwächter Dose, eines der vortreflichsten Auflösungsmittel. Es befördert das Sputum in der Peripneumonie, und ist in allen Fällen, wo Obstructiones viscerum sind, mit dem besten Erfolge zu gebrauchen. Bei der Manie muß man oft zu acht bis zehn Unzen täglich geben, ehe es irgend eine Ausleerung macht. Aber man

Fig. 4.18 (continued)



types that are made and how they're used. I simply want to make you aware that there's a 32-bit platform sitting under a 16-bit DOS facade, and that we can uncover it with the use of a DOS extender.

Let's begin by taking a look at DOS as an operating system and see what the problem is with it. DOS is a 16-bit operating system that was designed in the 1970s. (That's right: the 1970s. DOS is really a conversion of CP/M.) DOS was originally designed to operate on the Intel 8088 in the IBM PC. The operating system had been designed to access one megabyte of RAM, had an interrupt structure, was open-ended, and (for its time) was a great breakthrough in technology. Alas, time marched on and newer versions of the 8088 were designed. The 8086 came out and was a full, 16-bit version of the 8088. This didn't make much of a difference to DOS because the 8086 could only address one megabyte, and was relatively the same.

Then Intel came out with the 80286, and this is where things started changing. The 80286 could directly address a total of 16 megabytes. This was unheard of! At the time, the total power of the new 286 hadn't yet been realized. It had extra instructions, hardware for multitasking, and could operate in a new mode called *protected mode*. Then, too, the cost of memory was decreasing, and more and more PCs were fitted with the full one megabyte of RAM a PC could access.

This was the point when DOS should have been put to rest and DOS II should have been created as the next generation. However, as we all know, this isn't what happened. Intel started creating processors (like they were cheese sandwiches) that were extremely advanced. These new processors—the 386, 486, and now the Pentium—could access gigabytes of RAM and had all kinds of added hardware for internal caching, multitasking, and so forth. However, DOS stayed more or less its 16-bit self. Even though the processors that began with the 386 had enhanced capabilities, DOS didn't answer these challenges—so other manufacturers did. Beginning with the introduction of the 80286, companies started creating what were called *DOS extenders*. These were mini-operating systems that could place the processor into what was called *protected mode*. This new mode of the processor allowed it to use more instructions and address its full memory space.

When the first DOS extenders came out for the 80286, the performance increase wasn't that noticeable. The added memory addressing ability was, though. At this point DOS programs could access whatever amount of memory was in the PC and do it in a straightforward way. The difference in performance didn't arise until the release of the 80386. Here was truly a next-generation processor with all kinds of features. However, when the 80386 was plugged into a PC and run under DOS, it was just a fast 8086 running in what was called *real mode* or *8086 emulation mode*. The 80386 was a full, 32-bit processor that was being held back by DOS.

457

Fig. 4.18 Examples of document images captured by (**a**) flatbed scanner and (**b**) digital camera

model with cubic B-splines, Wu and Agam [75] use a nonlinear curve for each text line, and L. Zhang and Tan [79] represent the text lines using natural cubic splines. The above mentioned techniques suffer from various limitations. Specifically, the approach in [37] is not efficient in the case of inhomogeneous line spacing and method [75] is based on several heuristics, while it requires that the user should interactively specify the four corner points of the warped image which is not practical and cannot handle nonuniform columns in the target mesh, as well. In [79], L. Zhang and Tan assume that the book spine is found along iso-parametric lines. In a recent work of Stamatopoulos et al. [63], a goal-oriented rectification methodology to compensate for undesirable document image distortions aiming to improve the OCR result is presented. The approach relies upon a coarse-to-fine strategy. It first detects words and text lines to rectify the document image in a coarse scale and then further normalize individual words in finer detail using baseline correction. Coarse rectification is

Table 4.6 Overview of key document image dewarping techniques

Reference	Category	Short description	Remarks
Cao et al. [12]	3-D document shape	Rectifies warping distortions in document images by constructing a cylinder model. The skeleton of each horizontal text line is extracted to help estimate the parameter of the model and rectify the images	It does not require any information about the image formation process or any specific stereographic device
Liang et al. [42]	3-D document shape	It models the page surface by curved developable surfaces to estimate the 3-D shape of the page using texture flow fields	The document has to be either flat or smoothly curved and the camera a standard pinhole camera
Zhang et al. [80]	3-D document shape	It is based on a shape-from-shading formulation in order to reconstruct the 3-D shape of the document's surface	Does not produce satisfactory results in some cases of images with degradations
Lavialle et al. [37]	2-D document image processing	It is based on the detection of distorted text lines at the original document image using an analytical model with cubic B-splines	Not efficient in the case of inhomogeneous line spacing
Wu and Agam [75]	2-D document image processing	It is based on the detection of distorted text lines using a nonlinear curve for each text line	Based on several heuristics and requires user interaction
L. Zhang and Tan [79]	2-D document image processing	It represents the text lines using natural cubic splines	Assumes that the book spine is found along iso-parametric lines
Stamatopoulos et al. [63]	2-D document image processing	First detects words and text lines to rectify the document image in a coarse scale and then further normalize individual words in finer detail using baseline correction	A goal-oriented rectification aiming to improve the OCR result
Ulges et al. [69]	2-D document image processing	Estimates quadrilateral cell for each letter based on local baseline finding and then maps to a rectangle of corrected size and position in the dewarped image	Only for documents with straight lines that are approximately equally spaced and sized and the spacing between words is not large
Lu et al. [46]	2-D document image processing	Divides images into multiple quadrilateral patches based on the exploitation of the vertical stroke boundaries and text baselines	Based on several heuristics and limited on documents printed in Latin-based languages
Brown and Tsoi [10]	2-D document image processing	It uses document boundary interpolation to correct geometric distortions and shading artifacts present in images of art-like materials	Limited to some specific document distortions
Masalovitch and Mestetskiy [48]	2-D document image processing	Approximates deformation of interlinear spaces of image based on elements of image's skeleton that lie between the text lines	Sensitive to the approximation of vertical border deformation in text blocks

accomplished with the aid of a computationally low-cost transformation which addresses the projection of a curved surface to a 2-D rectangular area.

A few more rectification techniques also rely on text line detection but, in contrast with the above mentioned techniques, they emphasize on baseline finding. Ulges et al. [69] estimate quadrilateral cell for each letter based on local baseline finding and then map to a rectangle of corrected size and position in the dewarped image. Their method is not generic since it is based on the assumption that the original page contains only straight lines that are approximately equally spaced and sized and spacing between words is not large. Lu et al. [46] restore the document by dividing images into multiple quadrilateral patches based on the exploitation of the vertical stroke boundaries (VSBs) and text baselines. This method is based on several heuristics and is limited on documents printed in Latin-based languages.

There are also rectification techniques that do not rely on the detection of distorted text lines but they aim to find spatial transformations between the warped and dewarped document images by analyzing the 2-D content such as document boundaries or known reference points. Brown and Tsoi [10] use document boundary interpolation to correct geometric distortions and shading artifacts present in images of art-like materials. They use a physical pattern to guide the uniform parameterization, so it is limited to some specific document distortions. Masalovitch and Mestetskiy [48] approximate deformation of interlinear spaces of image based on elements of image's skeleton that lie between the text lines. This method is sensitive to the approximation of vertical border deformation in text blocks, which diminish accuracy.

Conclusion

In this chapter, we introduced basic image processing algorithms used in document image analysis while the focus was given on the techniques used for document image binarization, enhancement, and normalization. Document image binarization is used to separate the text from the background regions. Document image enhancement aims to improve the quality of document images by diminishing artifacts such as low contrast and uneven background illumination, bleed-through and shadow effects, damaged characters, and noisy black borders. Document image normalization refers to the task of restoring the document image horizontal alignment after correcting possible page skew, character slant, warping, and perspective distortions. For all these tasks, we categorized the main approaches while we presented an overview of key techniques using comparative tables.

Cross-References

►Document Creation, Image Acquisition and Document Quality
►Handprinted Character and Word Recognition
►Page Segmentation Techniques in Document Analysis
►Text Segmentation for Document Recognition

References

1. Akiyama T, Hagita N (1990) Automated entry system for printed documents. Pattern Recognit 23(11):1141–1154
2. Amin A, Fischer S (2000) A document skew detection method using the Hough transform. Pattern Anal Appl 3(3):243–253
3. Aradhye HB (2005) A generic method for determining up/down orientation of text in roman and non-roman scripts. Pattern Recognit 38:2114–2131
4. Avila BT, Lins RD (2004) A new algorithm for removing noisy borders from monochromatic documents. In: ACM-SAC'2004, Cyprus, Mar 2004. ACM, pp 1219–1225
5. Avila BT, Lins RD (2004) Efficient removal of noisy borders from monochromatic documents. In: ICIAR 2004, Porto, Portugal. LNCS 3212, pp 249–256
6. Badekas E, Papamarkos N (2007) Optimal combination of document binarization techniques using a self-organizing map neural network. Eng Appl Artif Intell (Elsevier) 20:11–24
7. Badekas E, Nikolaou N, Papamarkos N (2006) Text binarization in color documents. Int J Imaging Syst Technol 16(6):262–274
8. Baird HS (1987) The skew angle of printed documents. In: SPSE 40th conference and symposium on hybrid imaging systems, Rochester, pp 21–24
9. Bozinovich A, Srihari A (1989) Off-line cursive script word recognition. Trans Pattern Anal Mach Intell II(1):69–82
10. Brown MS, Tsoi YC (2006) Geometric and shading correction for images of printed materials using boundary. IEEE Trans Image Process 15(6):1544–1554
11. Cao Y, Li H (2003) Skew detection and correction in document images based on straight-line fitting. Pattern Recognit Lett 24(12):1871—1879
12. Cao H, Ding X, Liu C (2003) Rectifying the bound document image captured by the camera: a model based approach. In: 7th international conference on document analysis and recognition, Edinburgh, pp 71–75
13. Caprari RS (2000) Algorithm for text page up/down orientation determination. Pattern Recognit Lett 21:311–317
14. Chen Y, Leedham G (2005) Decompose algorithm for thresholding degraded historical document images. IEE Vis Image Signal Process 152(6):702–714
15. Chen YK, Wang JF (2001) Skew detection and reconstruction of color-printed document images. IEICE Trans Inf Syst E84-D(8):1018–1024
16. Cheriet M, Said JN, Suen CY (1998) A recursive thresholding technique for image segmentation. IEEE Trans Image Process 7(6):918–921
17. Chou CH, Chu SY, Chang F (2007) Estimation of skew angles for scanned documents based on piecewise covering by parallelograms. Pattern Recognit 40:443–455
18. Chou C-H, Lin W-H, Chang F (2010) A binarization method with learning-built rules for document images produced by cameras. Pattern Recognit 43(4):1518–1530
19. Ding Y, Kimura F, Miyake Y (2000) Slant estimation for handwritten words by directionally refined chain code. In: 7th international workshop on Frontiers in handwriting recognition (IWFHR 2000), Amsterdam, pp 53–62
20. Ding Y, Ohyama W, Kimura F, Shridhar M (2004) Local Slant estimation for handwritten English words. In: 9th international workshop on Frontiers in handwriting recognition (IWFHR 2004), Tokyo, Japan, pp 328–333,
21. Drira F, LeBourgeois F, Emptoz H (2011) A new PDE-based approach for singularity-preserving regularization: application to degraded characters restoration. Int J Doc Anal Recognit. doi:10.1007/s10032-011-0165-5
22. Dubois E, Pathak A (2001) Reduction of bleed-through in scanned manuscript documents. In: Proceedings of the image processing, image quality, image capture systems conference, Apr 2001, pp 177–180
23. Fadoua D, Le Bourgeois F, Emptoz H (2006) Restoring ink bleed-through degraded document images using a recursive unsupervised classification technique. In: 7th international workshop on document analysis systems, (DAS 2006), Nelson, New Zealand, pp 38–49

24. Fan KC, Wang YK, Lay TR (2002) Marginal noise removal of document images. Pattern Recognit 35(11):2593–2611
25. Fan H, Zhu L, Tang Y (2010) Skew detection in document images based on rectangular active contour. Int J Doc Anal Recognit 13(4):261–269
26. Gatos B, Papamarkos N, Chamzas C (1997) Skew detection and text line position determination in digitized documents. Pattern Recognit 30(9):1505–1519
27. Gatos B, Pratikakis I, Perantonis SJ (2006) Adaptive degraded document image binarization. Pattern Recognit 39:317–327
28. Gatos B, Pratikakis I, Perantonis SJ (2008) Efficient binarization of historical and degraded document images. In: 8th international workshop on document analysis systems (DAS'08), Nara, Sept 2008, pp 447–454
29. Haji MM, Bui TD, Suen CY (2009) Simultaneous document margin removal and skew correction based on corner detection in projection profiles. In: ICIAP 2009, Solerno, Italy. LNCS 5716, pp 1025–1034
30. Huang S, Ahmadi M, Sid-Ahmed MA (2008) A hidden Markov model-based character extraction method. Pattern Recognit 41(9):2890–2900
31. Hull JJ (1998) Document image skew detection: survey and annotated bibliography. In: Hull JJ, Taylor SL (eds) Document analysis systems II. World Scientific, Singapore/River Edge, pp 40–64
32. Kamel M, Zhao A (1993) Extraction of binary character/graphics images from grayscale document images. Comput Vis Graph Image Process 55:203–217
33. Kavallieratou E, Fakotakis N, Kokkinakis G (2001) Slant estimation algorithm for OCR systems. Pattern Recognit 34:2515–2522
34. Kim G, Govindaraju V (1997) A Lexicon driven approach to handwritten word recognition for real-time applications. IEEE Trans Pattern Anal Mach Intell 19(4):366–379
35. Kim I-K, Jung D-W, Park R-H (2002) Document image binarization based on topographic analysis using a water flow model. Pattern Recognit 35:265–277
36. Kimura F, Shridhar M, Chen Z (1993) Improvements of a lexicon directed algorithm for recognition of unconstrained handwritten words. In: 2nd international conference on document analysis and recognition (ICDAR 1993), Tsukuba City, Japan. Oct 1993, pp 18–22
37. Lavialle O, Molines X, Angella F, Baylou P (2001) Active contours network to straighten distorted text lines. In: International conference on image processing, Thessaloniki, pp 748–751
38. Le DX, Thoma GR (1996) Automated borders detection and adaptive segmentation for binary document images. In: International conference on pattern recognition, (ICPR 1996), Vienna, Austria, p III: 737–741
39. Le DS, Thoma GR, Wechsler H (1994) Automated page orientation and skew angle detection for binary document images. Pattern Recognit 27(10):1325–1344
40. Leung CC, Chan KS, Chan HM, Tsui WK (2005) A new approach for image enhancement applied to low-contrast–low-illumination IC and document images. Pattern Recognit Lett 26:769–778
41. Li S, Shen Q, Sun J (2007) Skew detection using wavelet decomposition and projection profile analysis. Pattern Recognit Lett 28:555–562
42. Liang J, DeMenthon D, Doermann D (2008) Geometric rectification of camera-captured document images. IEEE Trans PAMI 30(4):591–605
43. Likforman-Sulem L, Darbon J, Barney Smith EH (2011) Enhancement of historical printed document images by combining total variation regularization and Non-local Means filtering. Image Vis Comput J 29(5):351–363
44. Liu H, Wua Q, Zha H, Liu X (2008) Skew detection for complex document images using robust borderlines in both text and non-text regions. Pattern Recognit Lett 29:1893–1900
45. Lu Y, Tan CL (2003) A nearest-neighbor chain based approach to skew estimation in document images. Pattern Recognit Lett 24:2315–2323
46. Lu SJ, Chen BM, Ko CC (2006) A partition approach for the restoration of camera images of planar and curled document. Image Vis Comput 24(8):837–848

47. Makridis M, Nikolaou N, Papamarkos N (2010) An adaptive technique for global and local skew correction in color documents. Expert Syst Appl 37(10):6832–6843
48. Masalovitch A, Mestetskiy L (2007) Usage of continuous skeletal image representation for document images de-warping. In: International workshop on camera-based document analysis and recognition, Curitiba, pp 45–53
49. Moghaddam RF, Cheriet M (2010) A variational approach to degraded document enhancement. IEEE Trans Pattern Anal Mach Intell 32(8):1347–1361
50. Niblack W (1986) An introduction to digital image processing. Prentice-Hall, Englewood Cliffs, pp 115–116
51. Nomura S, Yamanaka K, Shiose T, Kawakami H, Katai O (2009) Morphological preprocessing method to thresholding degraded word images. Pattern Recognit Lett 30(8):729–744
52. Obafemi-Ajayi T, Agam G, Frieder O (2010) Historical document enhancement using LUT classification. Int J Doc Anal Recognit 13:1–17
53. O'Gorman L (1993) The document spectrum for page layout analysis. IEEE Trans Pattern Anal Mach Intell 15(11):1162–1173
54. Otsu N (1979) A threshold selection method from Gray-level histograms. IEEE Trans Syst Man Cybern 9(1):377–393
55. Pratt WK (2012) Digital image processing. Wiley, New York
56. Sauvola J, Pietikainen M (2000) Adaptive document image binarization. Pattern Recognit 33:225–236
57. Sezgin M, Sankur B (2004) Survey over image thresholding techniques and quantitative performance evaluation. J Electron Imaging 13(1):146–165
58. Shafait F, Breuel TM (2009) A simple and effective approach for border noise removal from document images. In: 13th IEEE international multi-topic conference, Islamabad, Dec 2009, pp 1–5
59. Shafait F, van Beusekom J, Keysers D, Breuel TM (2008) Document cleanup using page frame detection. Int J Doc Anal Recognit 11:81–96
60. Singh C, Bhatia N, Kaur A (2008) Hough transform based fast skew detection and accurate skew correction methods. Pattern Recognit 41(12):3528–3546
61. Solihin Y, Leedham C (1999) Integral ratio: a new class of global thresholding techniques for handwriting images. IEEE Trans Pattern Anal Mach Intell 21:761–768
62. Stamatopoulos N, Gatos B, Kesidis A (2007) Automatic borders detection of camera document images. In: 2nd international workshop on camera-based document analysis and recognition (CBDAR'07), Curitiba, Sept 2007, pp 71–78
63. Stamatopoulos N, Gatos B, Pratikakis I, Perantonis SJ (2011) Goal-oriented rectification of camera-based document images. IEEE Trans Image Process 20(4):910–920
64. Su B, Lu S, Tan CL (2011) Combination of document image binarization techniques. In: 11th international conference on document analysis and recognition, Beijing, 18–21 Sept 2011
65. Tonazzini A (2010) Color space transformations for analysis and enhancement of ancient degraded manuscripts. Pattern Recognit Image Anal 20(3):404–417
66. Trier ØD, Taxt T (1995) Evaluation of binarization methods for document images for document images. IEEE Trans Pattern Anal Mach Intell 17(3):312–315
67. Tsai CM, Lee HJ (2002) Binarization of color document images via luminance and saturation color features. IEEE Trans Image Process 11(4):434–451
68. Tseng YH, Lee HJ (2008) Document image binarization by two-stage block extraction and background intensity determination. Pattern Anal Appl 11:33–44
69. Ulges A, Lampert CH, Breuel TM (2005) Document image dewarping using robust estimation of curled text lines. In: 8th international conference on document analysis and recognition, Seoul, pp 1001–1005
70. van Beusekom J, Shafait F, Breuel TM (2010) Combined orientation and skew detection using geometric text-line modeling. Int J Doc Anal Recognit 13(2):79–92
71. Vinciarelli A, Luettin J (2001) A new normalization technique for cursive handwritten words. Pattern Recognit Lett 22(9):1043–1050

72. Vonikakis V, Andreadis I, Papamarkos N (2011) Robust document binarization with OFF center-surround cells. Pattern Anal Appl 14:219–234
73. Wahl FM, Wong KY, Casey RG (1982) Block segmentation and text extraction in mixed text/image documents. Comput Graph Image Process 20:375–390
74. Wang B, Li XF, Liu F, Hu FQ (2005) Color text image binarization based on binary texture analysis. Pattern Recognit Lett 26:1650–1657
75. Wu C, Agam G (2002) Document image De-warping for text/graphics recognition. In: Joint IAPR international workshop on structural, syntactic and statistical pattern recognition, Windsor, pp 348–357
76. Yan H (1993) Skew correction of document images using interline cross-correlation. Graph Models Image Process 55(6):538–543
77. Yang Y, Yan H (2000) An adaptive logical method for binarization of degraded document images. Pattern Recognit 33:787–807
78. Yin PY (2001) Skew detection and block classification of printed documents. Image Vis Comput 19(8):567–579
79. Zhang L, Tan CL (2005) Warped image restoration with applications to digital libraries. In: 8th international conference on document analysis and recognition, Seoul, pp 192–196
80. Zhang L, Yip AM, Brown MS, Tan CL (2009) A unified framework for document restoration using in painting and shape-from-shading. Pattern Recognit J 42(11):2961–2978

Further Reading

Gonzalez RC, Woods RE (2007) Digital image processing, 3rd edn
Petrou M, Petrou C (2010) Image processing: the fundamentals, 2nd edn
Pratt WK (2012) Digital Image Processing