Page Segmentation and Zone Classification: The State of the Art

Article · November 1999

CITATIONS
45

READS
927

3 authors, including:

Oleg Okun
Cognizant Solutions GmbH, Hamburg
71 PUBLICATIONS **1,263** CITATIONS
SEE PROFILE

Matti Pietikainen
University of Oulu
76 PUBLICATIONS **3,651** CITATIONS
SEE PROFILE

LAMP-TR-036
CAR-TR-927
CS-TR-4079

MDA9049-6C-1250
November 1999

Page Segmentation and Zone Classification: The State of the Art

Oleg Okun[1] David Doermann[2], and Matti Pietikäinen[1,*]

[1]Machine Vision and Media Processing Unit
Infotech Oulu and Dept. of EE, University of Oulu
P.O.Box 4500, FIN-90401 Oulu, Finland

[2]Language and Media Processing Laboratory
Institute for Advanced Computer Studies
University of Maryland
College Park, MD 20742-3275

Abstract

Page segmentation and zone classification are key areas of research in document image processing, because they occupy an intermediate position between document preprocessing and higher-level document understanding such as logical page analysis and OCR. Such analysis of the page relies heavily on an appropriate document model and results in a representation of the physical structure of the document. The purpose of this review is to analyze progress made in page segmentation and zone classification and suggest what needs to be done to advance the field.

Keywords: Page segmentation; Zone classification; Document layout

The support of this research by the Department of Defense under contract MDA 9049-6C-1250 and by the Technology Development Center (TEKES) of Finland is gratefully acknowledged.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
NOV 1999		**00-11-1999 to 00-11-1999**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Page Segmentation and Zone Classification: The State of the Art	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Language and Media Processing Laboratory,Institute for Advanced Computer Studies,University of Maryland,College Park,MD,20742-3275**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**		**37**	

LAMP-TR-036
CAR-TR-927
CS-TR-4079

November 1999

Page Segmentation and Zone Classification: The State of the Art

Oleg Okun, David Doermann, and Matti Pietikäinen

Page Segmentation and Zone Classification: The State of the Art

Oleg Okun[1] David Doermann[2], and Matti Pietikäinen[1,*]

[1]Machine Vision and Media Processing Unit
Infotech Oulu and Dept. of EE, University of Oulu
P.O.Box 4500, FIN-90401 Oulu, Finland

[2]Language and Media Processing Laboratory
Institute for Advanced Computer Studies
University of Maryland
College Park, MD 20742-3275

Abstract

Page segmentation and zone classification are key areas of research in document image processing, because they occupy an intermediate position between document preprocessing and higher-level document understanding such as logical page analysis and OCR. Such analysis of the page relies heavily on an appropriate document model and results in a representation of the physical structure of the document. The purpose of this review is to analyze progress made in page segmentation and zone classification and suggest what needs to be done to advance the field.

Keywords: Page segmentation; Zone classification; Document layout

The support of this research by the Department of Defense under contract MDA 9049-6C-1250 and by the Technology Development Center (TEKES) of Finland is gratefully acknowledged.

1 Introduction

Paper-based documents contain information in various forms such as text, graphics, pictures, mathematical formulas, and tables. To fully "understand" each type of information, it is necessary to apply domain-specific analysis techniques specific to that document type. Initially, however, a scanned image of the document must be divided or segmented into homogeneous regions and each region should be classified so that appropriate analysis can be applied—for example, so that graphics can be vectorized, pictures can be compressed, and text can be segmented into lines, words, and/or characters and recognized. Segmentation and classification are therefore of great importance for document image processing and its applications, because they define a baseline for the whole process of information conversion to digital form.

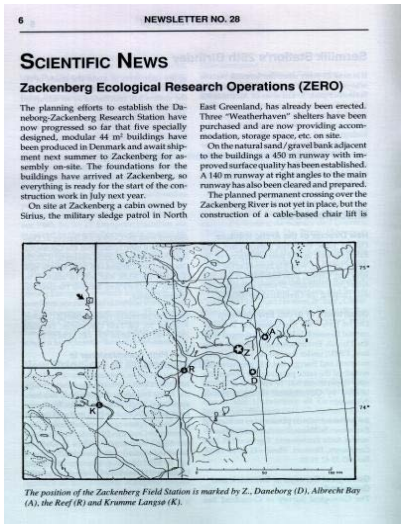
A great deal of work has been done on page segmentation and zone classification and various methods have been proposed (see the earlier surveys [36, 89]). The purpose of this paper is to survey existing methods, to highlight special features, and to suggest tasks that will lead to better results in future applications. This work does not describe the methods in detail, but rather assumes that a brief and simple introduction to the subject is often more helpful than reading many papers and trying to understand the authors' thoughts. Here, we focus on geometrical layout analysis without considering logical layout analysis. Topics such as the division of pages into headers, footers, title, and abstracts, and how text regions are related to each other, are not directly within the scope of this paper.

We will review page segmentation and zone classification methods for document images consisting of text, binary graphics, and binarized, halftone or color pictures. Several examples of such images are shown in Fig. 1.

This paper has the following structure. Section 2 describes the basic elements that typically appear in document images and common image types. Section 3 briefly considers various classes of document images, tasks specific for each class, and known difficulties in their layout analysis. Section 4 presents the state of the art of page segmentation and zone classification methods. Section 5 presents our view on the state of the art and overviews related document image analysis tasks such as document compression, representation and benchmarking of document layout analysis algorithms. Section 6 concludes the paper.

(a) Advertisement

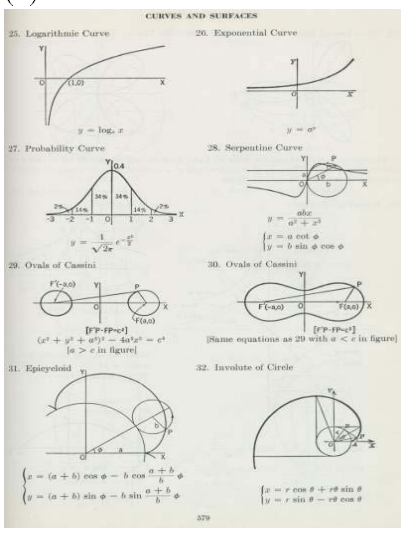
(b) Advertisement



(c) Newsletter

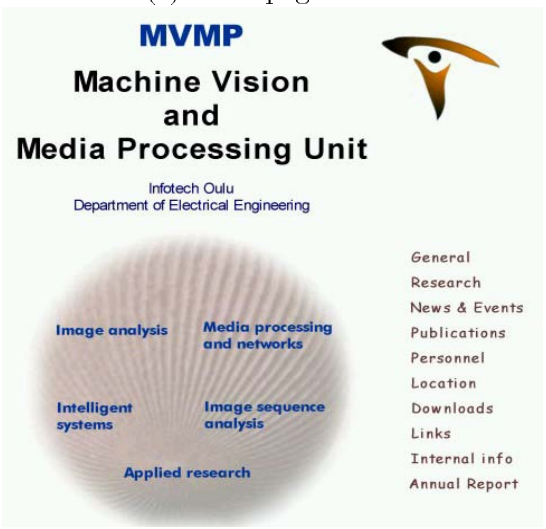
(d) Article

(e) Article



(f) Book page

(g) WWW page



(h) WWW page

Figure 1: Examples of document images

2 Basic characterization of document images

Documents are usually scanned and represented as binary (2 bits per pixel), gray-scale (typically 8 bits per pixel), or color (typically 8–24 bits per pixel) images. The type of image used is application-dependent and transformations between types, such as color to gray-scale (when a color image is split into three grey-scale images corresponding to three color planes (R,G,B)); color to binary (when applying edge detection directly to a color image); and gray-scale to binary (when using binarization), are often done in order to speed up computations.

Basic elements or entities that can be present in document images include but are not limited to:

- Text (characters and digits in the text body, titles, headings, cells of tables, figure captions) nested in pictures and graphics as annotations.
- Tables (with and without ruling lines as field separators)
- Mathematical expressions
- Binarized, halftone, and color pictures
- Graphics (flow charts, line drawings, plots, diagrams, logos, etc.).

For some regions that contain text, it is often unclear whether to classify them as text or as non-text. For example, we may wish to treat the entire region as non-text, if the text is nested in or semantically close to pictures or graphics. On the other hand, it can be useful to process regions as text when this explains, for example, a line drawing. In most cases, even if text is initially classified as non-text, it can be extracted as part of zone-specific processing.

Geometric document layout and *skew* are two important aspects of a document image. A document's layout is the way in which document elements (text, pictures, graphics, etc.) are arranged on the document image. There are three basic types of layouts: Manhattan layout, where regions are constrained polygons whose boundaries are straight horizontal and vertical lines; rectangular layout, which is a specific case of Manhattan layout; and arbitrary layout, where boundaries form unconstrained polygons or overlapping regions. Document skew is the slant of the document image with respect to the primary orientation of the page. Skew is most often introduced by improper positioning of the document on a scanning device. In many

cases, this negatively influences the performance of page segmentation and zone classification methods.

Most layout analysis methods segment and classify the image into three basic classes: text, graphics, and pictures. Mathematical expressions may be initially classified as text, and tables may be classified as graphics. When doing this, it is assumed that a finer classification will be done at the document image understanding stage, where the text is divided into logical components.

3 Document classes and applications

It is highly unlikely that a single generic method will be developed that can process all classes of documents because of their variety and complexity, although a number of methods designed to analyze the images of several different classes of documents have been proposed. It appears most useful to begin with classification of documents by their geometrical layouts so that when new applications arise, we can evaluate what layout analysis methods can be employed based on properties of the target class. Although we do not consider the following list to be comprehensive, nor do we consider each class to be exclusive, this taxonomy of document genres adequately highlights the field:

1. structured articles (in journals, newspapers, and newsletters),
2. documents with unconstrained layout (advertisements, cover and title pages of CDs, books, and journals),
3. semi-structured layouts (envelopes, post and business cards, bank checks, forms, and table-like documents),
4. maps and engineering drawings,
5. non-traditional documents (WWW pages and video frames).

In this section, we will give a brief overview of the main segmentation and classification tasks to be addressed for instances of each class mentioned above together with known problems.

3.1 Structured articles

Journals, newspapers and newsletters are a primary source of information and are widely published and easily accessible. Their images can contain zones of many types (text of different

font styles/sizes, various types of graphics, binarized, halftone, and color pictures) and can be scanned as binary, gray-scale or color. Zones are typically physically separated from each other; however, the gaps between them can be very narrow. In some cases, text can be embedded into graphics or pictures, and can be either darker (normal printing) or lighter (inverse printing) than the background. Articles in journals, newspapers, and newsletters are examples of structured documents, because all elements on a page are typically ordered and linked together based on general rules. Examples of such rules include: the reading order of text blocks is from top to bottom, left to right; figure captions are located after the corresponding figures; text data form relatively large regions; figures or tables appear only after references to them in the text. Newspapers have more complex and less structured logical organization than journals or newsletters. They can be typically described with the Manhattan layout but can also have arbitrary (unconstrained) layout. Cover pages of technical journals may have a table-like structure, but they are typically weakly constrained for non-technical magazines. Although these models can be language- or even publication-specific, they are typically quantifiable. The main tasks to be addressed include skew estimation and correction (sometimes optional), document segmentation into homogeneous regions, and classification of these regions as background, text, graphics, and pictures (see, for example, [7, 20, 27, 30, 34, 53] for recent work; other papers are cited in Section 4). Text lines with no skew are typically horizontal or vertical, but they are slanted with respect to the X- or Y-axis if skew is present. Usually all regions on the image have the same skew (global skew), but cases where some regions have different orientations than others are also possible.

Despite the rather structured layout in many cases, the major problems are due to different image types, application conditions that must be satisfied (skew-, layout-, script-independence), and noise.

3.2 Unconstrained layouts

The term "unconstrained layout" reflects a lack of general rules when defining the documents of this class. More generally, the layout of such documents depends on their designer's goal.

3.2.1 Advertisements

Advertisements are usually printed in magazines or newspapers, but they can also be placed on the Web as electronic documents. Their layout is typically more arbitrary than that of journal

and newspaper articles. Multiple skew for text and non-text regions, curved text lines whose characters are not aligned along a straight line, and text nested in or touching pictures often occur in order to emphasize important information and to attract the reader's attention [42, 48]. In short, advertisements are unstructured documents.

3.2.2 Cover and title pages

The images of these documents are often in color and they can contain arbitrarily oriented text of a large variety of a priori unknown font sizes printed on a complex color or textured background.

Layout analysis of images of this class requires text extraction and recognition [17, 37, 48, 68, 84, 108, 109], which is useful for information retrieval. Text queries are easier to formulate than ones described by non-text features such as shape, texture, and color.

The challenges are due to arbitrary document layout and complex color background, which makes accurate text detection difficult.

3.3 Semi-structured layouts

Semi-structured layouts have structural elements in which a user should enter information or which confine the entered data, but there are typically no limitations on the locations of these elements. This means that two documents belonging to this class can have different layouts, although they may both contain the same structural elements.

3.3.1 Postal envelopes and bank checks

Postal correspondence and bank checks are examples of semi-structured documents since they have predefined sets of fields such as the address block or the amount of money, but the locations of these fields within the document are subject only to convention. Business cards are also included in this class because they have a similar logical structure. The images of this class of documents may be binary, gray-scale or color and text may be printed on a complex textured background.

The main tasks deal with the identification of different fields which are specific to given types of documents. For bank checks, they may include: signature, check number, date, courtesy amount (amount of payment in a numeric format), legal amount (amount of payment in a character format), account number, payee's name, address of financial institution, and/or

logos [1, 18, 25, 40, 54, 57, 85, 105]. For business cards, they may include holder's name, affiliation, and address [19, 97]. For postal correspondence, the address block or blocks, stamp, bar code, and postage paid indications should be identified [22, 69, 92, 96, 99, 101, 104]. In systems such as that used in the British Postal Office, a stamp value is also read for revenue protection [69].

The major problems result from 1) a complex background, making it difficult to binarize or segment the image, 2) a mixture of printed and handwritten text touching or intersecting the ruling lines, 3) changes in illumination, 4) arbitrary document orientation on a moving platform, and 5) restrictions on processing time.

3.3.2 Forms and table-like documents

Forms and table-like documents such as questionnaires or invoices consist of fields or cells in which handwritten or printed data should be entered. The initial images are usually binary, although color forms exist too. In the latter case, however, the image background is uniform so that it does not seem to be very difficult to separate text from it. In addition to text, invoices can also contain small pictures such as logos. Forms and table-like documents are semi-structured documents with limitations imposed by different separators on the location of text data.

The basic tasks to be solved here are field isolation and removal of response or bounding boxes or/and ruling lines to extract the text inside each field [9, 14, 21, 38, 43, 61, 67, 77, 82, 90, 100, 107]. Invoices often also include the company's address and bank account location [13, 15, 55], making processing them somewhat similar to processing of bank checks.

There are two types of form analysis. The first is known form analysis, where a given form belongs to one of a set of known classes and a blank template (model) with empty fields or cells is available for it. In this case, the first task is to identify the form by using specific labels or structure that can be extracted from the model. Once the form type is identified, the locations of data fields in terms of their coordinates become known.

The second type is unknown form analysis, where no prior knowledge about the form type is used and an algorithm extracts a form structure based on separators between cells or fields. The second type is more flexible, but it may be more computationally expensive and sometimes less accurate.

Problems arise when the text data are written outside predefined boxes or fields; this makes

it difficult to locate them properly. Image transformations (rotation, scaling and translation) also complicate the process of matching a predefined model to the input image. For invoices sent through fax machines from sellers to customers, the corresponding images often have poor quality due to the lossy compression used in fax transmission. Stamps on text areas can also negatively affect text data extraction for such documents.

3.4 Maps and drawings

3.4.1 Maps

Map images usually require gray-scale or color to capture detail. Color maps consist of several color layers, where each layer corresponds to a particular type of information. Text data (names of countries, cities, lakes, ...) are embedded into or surrounded by graphics (roads, different kinds of textures, ...).

Text data on maps and drawings forms small and sparse groups arbitrarily mixed with graphic elements. Though there are some rules for such "mixtures", they are only valid for a narrow family of maps, which does not allow one to generalize them and to assume that their organization is structured. In fact, text may appear everywhere and text lines may have arbitrary orientation.

Although there is a wide variety of types of maps (cadastral, hydrographical, topographical, etc.), analysis of map images usually consists of three tasks [4, 28, 60, 65, 71, 76, 79, 87, 91, 94]: 1) text/graphics separation, 2) label assignment to characters and special symbols, and 3) graphics vectorization. For color images, color separation is often done during scanning by using a special scanner designed for this purpose. Characters and symbols are separated from graphics at each color layer by their sizes and they are fed to recognition modules, while graphics are thinned and vectorized.

The large variety of maps leads to a large variety of methods and systems, each of which is usually capable of processing only one map type. Among the problems are characters and symbols touching or intersecting graphics, slanted or curved text strings, variable gaps between adjacent characters belonging to the same string, different font sizes (though it is possible to know in advance the fonts used in map creation), and some graphical symbols that may be similar to characters.

3.4.2 Engineering drawings

Engineering drawings have some challenges in common with maps in that the documents of both classes contain small and sparse text data among larger graphic regions. Unlike maps, however the images of engineering drawings are typically binary or gray-scale, are more sparse, and have tighter geometric constraints. Geometrical layout analysis of engineering drawings [10, 2, 23, 41, 64, 66, 98, 106] requires text/graphics separation, and classification of graphic components as lines (dashed or hatched), arcs, dimensions, etc.

The diversity of types of engineering drawings usually means that a particular method can be only applied to one or a few different types. Text touching graphics and often poor quality of the original paper document are the main problems.

3.5 WWW page images and video frames

WWW pages and video frames containing text are a relatively new class of documents. They appear initially as electronic documents, unlike the others described previously. These documents are typically in color and their images often have low resolution in order to keep their sizes small for fast loading and aliased for perceptual clarity. Their main elements are text and pictures arbitrarily located within a document. Text in WWW pages can appear either in a character format or embedded in a bitmap. For video frames, text is often a part of the a rather complex image, appearing as either scene or graphic text [59]. Characters are often rendered at a resolution 3–4 times lower than the 200–300 dpi used for many other document classes, making detection and recognition difficult.

WWW pages and video frames can be considered unstructured documents because their layout is not limited to predefined rules, and in the case of video, it is often impossible to set such rules. The primary task here is to locate text embedded in the image for purposes of information retrieval and video indexing [11, 33, 39, 48, 58, 63, 73, 81, 108, 109].

The main problems are 1) the colors of text and background may be close to each other due to low image resolution, or the text may have low contrast with the background (for example, the text embedded in a complex textured background), so that thresholding will not separate text and background; 2) text components may fall on a curved line, or in the worst case, they may not be aligned at all (wave-like lines); 3) text components may have non-uniform color, and may be fragmented into several subparts, not because of noise but because of their design;

4) effects of motion in video sequences.

4 The state of the art of page segmentation and zone classification methods

This section presents the state of the art of page segmentation and zone classification methods. Among all the document classes mentioned in Section 3, we have chosen structured articles (in journals, newspapers, and newsletters), documents with unconstrained layout (advertisements, cover and title pages of CDs, books, and journals), and non-traditional documents such as WWW pages, because such documents can be described as *pages* (either paper or electronic) usually consisting of different *zones*. Unlike these, line drawings, maps, bank checks and forms do not have a zone-like structure since they contain a mixture of text and non-text elements, often touching or intersecting.

This section is divided into three subsections reviewing different segmentation and zone classification methods. Here, segmentation means image partitioning into homogeneous zones or areas each containing only one data type such as text or graphics. These zones, however, do not have class labels after document segmentation so zone classification is required to assign labels to them based on the values of selected features. Many methods do both segmentation and zone classification simultaneously. Each method will be only briefly described with no details on its implementation.

4.1 Document segmentation

A traditional taxonomy divides document segmentation approaches into bottom-up (data-driven), top-down (model-driven), and hybrid (intermediate between the bottom-up and top-down) methods. The top-down techniques are useful and fast when one knows particulars of a document's layout a priori. In this case, the processing begins with a whole page at the highest level that is then divided into smaller regions such as blocks, lines, words, and characters. The bottom-up algorithms start from pixels and group them into connected regions that are then combined into larger structures. They are more robust with respect to variations in document layout but often slow. The hybrid methods occupy an intermediate place between the previous classes; that is, they often try to combine the high speed of the top-down methods and the robustness of the bottom-up methods. It is also possible to classify segmentation methods into texture- and non-texture-based, because there are techniques that treat the document regions

as textures of different classes. See [72] for a survey of texture-based methods.

4.1.1 Bottom-up methods

The methods described in [24, 26, 42, 44, 46, 47, 53, 95] are based on component analysis of binary images and grouping of the components into characters, lines, and blocks by using closeness between adjacent components and their sizes. Smearing, nearest neighbor search, and Voronoi diagrams are the main grouping methods. The processing times vary widely, depending on the method. The methods in this group are often tolerant to skew (sometimes even multiple skew) and arbitrary document layouts. The skew can also be computed as a by-product of text line extraction.

If fast processing or generality are requirements, the choice of data representation is crucial for the methods in this group. In [95] all regions are represented in a hierarchical tree structure for processing various document types, such as forms or journals, containing English and Kanji characters. The block adjacency graph provides very fast processing in [46, 47].

Iterative connectivity analysis of pre-classified square blocks together with a number of carefully selected small masks forms the larger regions in [78]. Connectivity at the block level can dramatically reduce the processing time in comparison with pixel connectivity.

The method in [75] extracts information about the co-occurrence of pixel values within a 5×5 window centered at each pixel by scanning the image row by row. The nearest neighbor clustering technique is then used to merge the pixels into homogeneous regions. The advantage of this method is that it can process both binary and gray-scale images. It is not tolerant to document skew or arbitrary layout.

Approaches based on the background analysis of binary images are considered in [5, 7, 12, 70, 74]. The most advanced method [5, 7] employs so-called white tiles. It first creates a net of rectangles, each representing a widest rectangular area of white (background) space, and then traces through these rectangles to identify the region contours. The method uses a flexible data representation that can represent both Manhattan and arbitrary layouts without splitting a complex-shaped region. It does not require prior skew correction and can process locally skewed text regions with different orientations. It is also fast, because pixel-based processing is performed only once.

Morphological operations (opening, closing) are applied in [29, 30, 49] to group pre-classified pixels into larger regions. Image smearing using the Run Length Smoothing (RLS) method is

employed in [31, 32, 34, 56, 80]. This operation is similar to a directional morphological dilation. Some methods of this group [29, 30, 32] are tolerant to document skew or/and arbitrary layout, while others [31, 34, 49, 56] seem not to be.

Text extraction from complex color document images is treated in [17, 48, 68, 73, 108–110]. Many of these methods assume that characters have a uniform color and are arranged in horizontal lines, and the colors of text and background are well separated. The method in [17] processes images of technical journal covers. It consists of color quantization for reducing the number of colors followed by edge- and color-based segmentation. The method in [48] extracts text by using multivalued image processing; it can be applied to advertisement images, WWW images, CD and book cover images, and video frames. A multivalued image may be a binary image (advertisement), gray-scale image, pseudo-color image (WWW image in GIF format), or full-color image (video, book, or CD cover). It is decomposed into multiple foreground and background-complementary foreground images so that connected component analysis can be used to identify text components.

The method in [68] appears to have several significant advantages over the others: it does not depend on text line skew, and it can process correctly merged characters and even adjacent lines. The method first detects connected components in one or several binary images. The number of images depends on the image type (binary, gray-scale or color). For example, for binary images, the connected components are detected on two images that have a positive and negative text contrast with respect to the background. Text lines are extracted by means of a hierarchical divisive procedure employing a number of heuristics. Gray-scale images of book covers were chosen for demonstration of this method's performance, though it seems that it can treat other document classes as well.

The method in [73] analyzes WWW images. It first reduces the number of colors by color quantization followed by connected component detection for each remaining color. After that, character candidates are selected by thinning the connected components and by computing their width and height during this process. The authors suggest that characters are composed of a set of strokes that have relatively fixed width and ratio between width and height. Strokes meeting these criteria are assumed to belong to characters and are used as input for a potential field approach that groups adjacent characters into lines. Like [68], this method is able to detect text of arbitrary orientation and even curved text lines.

The paper [109] proposes two methods of text location in complex images of book and CD

covers and video frames. The first method processes color images and uses a color histogram to find a set of dominant colors in the image. Then connected components are extracted for each color and several heuristics utilizing the size, alignment, and proximity of text components are applied to obtain candidate characters. The second method (it is also described in [108]) works on gray-scale images and employs the difference in the spatial variance between text and background. This feature is higher for text lines than for background, so that text lines can be identified by two sharp changes in values of the spatial variance (from low to high and from high to low). The spatial variance is computed for each pixel over a local neighborhood of $1 \times N$ pixels in the horizontal direction. Then horizontal edges are detected on the variance image by a Canny edge detector, and they are further merged into longer lines. Pairs of adjacent lines with opposite orientations are grouped together by using simple heuristics in order to locate text bounding boxes. This method must be modified somewhat if it is applied to color images; color, not gray-scale edge detection has to be used. Both methods are quite robust to variations in font, color, and text size. A hybrid method combining the two previous methods is also presented. It works better in cases where neither method alone can locate text regions. The method in [110] first quantizes the color space of an input image into color classes using a Euclidean minimum spanning tree. Then the text-like connected components in each color class are identified and grouped into horizontal lines using a set of heuristics.

4.1.2 Top-down methods

A common feature of the top-down methods described below is global-to-local processing. However, this may mean several quite different things:

- processing starts from a whole image and then descends to smaller blocks [3, 35],
- a global transformation is applied to a whole image and then pixels of the transformed image are grouped into clusters [45],
- processing can go from a coarse to a fine image representation [16, 20], where large regions are first quickly extracted in a coarse representation and are then refined in a finer representation.

A projection profile analysis that counts the number of pixels along a given direction is one of the best known top-down techniques. It is applied to binary images with no skew and

recursively divides an image into smaller regions based on valleys in vertical/horizontal profile histograms, which correspond to visual separators between rectangular blocks. Pixel projection profiles are utilized in [3], while profiles of bounding boxes are employed in [35] to speed up processing.

A global image transformation using a filter bank of several orientation-selective 2-D Gabor filters is applied in [45] to a gray-scale image followed by pixel clustering on the transformed image with a squared-error algorithm to detect text and non-text (background, picture) regions. This method is tolerant to skew but very slow.

Document segmentation of gray-scale images based on four Gaussian pyramids each with four levels is introduced in [20]. The pyramids represent four feature maps, where the features are the average, variance, threshold, and median. Processing is done from low (less detailed) to high (more detailed) image resolution. This method is skew- and layout-independent.

Text area detection on a textured background is an issue in [16], where a texture-based approach is used that consists of feature extraction with Laws' masks, coarse classification of 8×8-pixel non-overlapping blocks as text, background, and "fuzzy" (boundary between text and background), and fine text segmentation of "fuzzy" blocks at the pixel level. Both coarse and fine segmentation rely on stationary HMMs using from 4 to 8 states. The advantage of HMMs over many neural network training procedures is that each model is trained independently of the others—that is, when the data of a new class is added, a new HMM is created and trained only on the samples of that class (the other HMMs do not have to be retrained). The method is not sensitive to text skew, document layout, or script type.

4.1.3 Hybrid and other methods

In [93] a bottom-up RLS is applied to a binary image to detect text lines and non-text data, followed by top-down recursive X-Y cuts (RXYC) that combine the separate text lines into blocks. This method is simple to implement and quite fast, but it can only analyze rectangular layouts and requires prior binarization to process gray-scale images.

Adaptive split-and-merge segmentation of gray-scale document images into homogeneous regions represented as leaves of a quadtree is developed in [62]. Splitting and merging are performed at the same time. If some region is inhomogeneous, it is split into four rectangular subregions by thresholding based on projection profiles. If two adjacent regions (they may or may not be at the same segmentation level) are homogeneous and their union is also homoge-

neous, they are merged. The mean value and variance of pixel intensities in each region are used for making decisions about merging and splitting. The document layout can be arbitrary, but prior skew estimation and correction is necessary. This method can be applied to images of magazines, bank checks with complex backgrounds, and table-like documents. A similar technique is used in [86].

The method in [84] extracts text from complex color images of book and journal covers by using a hybrid analysis. The top-down technique recursively splits the image into rectangular blocks, and splitting terminates when there are pixels of at least two different colors inside a given block. Homogeneous blocks are considered to belong to the uniform background, while non-homogeneous blocks containing at least two different colors correspond to text. The bottom-up technique detects homogeneous regions of arbitrary shape by utilizing a region growing method. The results of both techniques are combined in order to verify whether a given region is text or not by assuming that text is horizontally aligned.

A method that does not belong to either group is presented in [88]. Gray-scale image segmentation is based on the fractal signature, which is less for the background than for image blocks. The method is not iterative and all processing is done in one step. This approach can be applied to documents that have complex layouts.

4.2 Document zone classification

A large number of methods [3, 26, 42, 46, 48, 68, 95, 109, 110] uses features of connected components to separate text and non-text in binary images. These features include the sizes of a connected component and of its nearest neighbors, alignment, proximity, and elongation.

Texture run-length statistics based on the occurrences of the black and white pixels within each segmented region are employed in [80, 93]. Run-length statistics calculated for four directions (horizontal, vertical, left diagonal and right diagonal) and classification based on a decision tree are proposed in [83]. Textural features of white tiles are also used in [6].

In [74] the classification utilizes cross-correlation between adjacent scanlines and the dependence of its behavior on the interline distance. A combination of run-length features and cross-correlation between pixels is used in [78].

In [31, 32] the vertical projection profile separates text from non-text (it is periodic for text), while the black pixel distribution helps to discriminate graphics and pictures (it is sparse for the former and dense for the latter).

Gray level histograms computed at several levels of image resolution distinguish between background, text, graphics, and pictures in [20].

Classification using soft computation techniques is described in [29, 30, 49, 56]. Usually a neural network (multilayer Perceptron in many cases) is first trained on samples of text, graphics, and pictures and then is used for classification. Block sizes and run-length features are inputs to a neural network in [56]. Low-order moments of wavelet packet components, which are computed over small windows, are the features used for neural classification in [29, 30]. To make classification more reliable for each window, fuzzy integration of decisions obtained from neighboring windows is carried out. Decisions are integrated at several scales of image resolution and within each scale as well. In [49] a neural network learns a small set of masks which best discriminate between text, background, line drawings, and pictures. Convolving these masks with the input image produces texture features that are used for classification of each image pixel by the neural network into one of three classes (text+line drawings, halftone pictures, and background). The regions belonging to the first class are further binarized with a fixed global threshold and separated by the size of the connected component. This method is robust to different languages and can discriminate between the text of languages such as English and Chinese. The methods in [29, 30, 49] are different from the others, because window/pixel classification is performed *before* segmentation into regions.

4.3 Literature comparison

In this subsection, we present a comparison of the properties and performance of different page segmentation and zone classification methods. The results are given in Tables 1 and 2.

In Table 1, '**Ref.**' refers to a given method. '**Document type**' refers to the classes of document images processed by the method. '**Image type**', B, G, and C correspond to binary, gray-scale, and color images, respectively. '**Background**' can be uniform U (usually white or black), T (textured), or C (color). '**Layout**' can be R (rectangular), M (Manhattan), or A (arbitrary). '**Skew**' indicates whether a method is tolerant to some degree of skew (Yes) or not (No). '**Test set**' and '**AC %**' refer to the test set size and accuracy of a given method. The symbol '–' means that the given feature was not mentioned in the original paper. The symbol '+' indicates that the results are not described by a single digit, but several criteria are used such as fragmentation and over-merging rates. *It is worth pointing out that some methods can be more or less easily modified; their properties will then be changed.*

In Table 2, '**Time**' corresponds to the processing time in seconds on a computer '**Platform**', '**Image size**' refers to image size, while '**Res.**' stands for image resolution. Image sizes and resolutions are given in pixels/paper page format and dpi, respectively. The notations are the same as those in Table 1.

Table 1: The most important properties of document layout analysis methods.

Ref.	**Document type**	**Image type**	**Back-ground**	**Lay-out**	**Skew**	**Test set**	**AC %**
[3]	Journals, newspapers	B	U	R	No	33	94.8–97.2
[7]	Journals, newspapers, newsletters, advertisements	B	U	A	Yes	40	+
[16]	Unknown	G	T	A	Yes	–	–
[17]	Journal covers	C	C	A	No	100	95.2–98
[20]	Journals	G	U	A	Yes	100	–
[26]	Journals, business cards, technical reports	B	U	A	No	30	93–100
[30]	Journals	G	U	A	Yes	–	–
[31]	Journals	B	U	R	No	–	–
[32]	Journals	B	U	R	Yes	30	92–97
[34]	Newspapers	B	U	M	No	100	96
[35]	Journals	B	U	R	No	150	–
[42]	Advertisements, letters, envelopes	B	U	A	Yes	150	–
[45]	Newspapers	G	U	A	Yes	–	–
[46]	Journals	B	U	M	No	150	91.1–99.4

Table 1 continued

Ref.	Document type	Image type	Back-ground	Lay-out	Skew	Test set	AC %
[48]	Advertisements,	B,	U, C	A	No	26	99.2
	WWW images,	G,				54	97.6
	book covers,	C				30	72.0
	video					6,952	94.7
[49]	Journals	G	U	M	No	–	–
[53]	Journals, newspapers	B	U	A	Yes	114	+
[56]	Journals	B	U	R	No	50	98.18–99.61
[62]	Journals, forms, bank checks	G	U, T	A	No	–	–
[68]	Book covers	G	U	A	Yes	100	91.2
[70]	Journals	B	U	A	Yes	–	–
[73]	WWW pages	C	C	A	Yes	200	88.8–92
[74]	Journals	B	U	R	Yes	–	–
[75]	Journals	B, G	U	M	No	–	–
[78]	Journals	B	U	R	No	–	99–100
[80]	Journals, newspapers	B	U	R	No	100	–
[83]	Journals	B	U	R	No	979	97
[84]	Book and journal covers	C	C	A	No	16	–
[88]	Journals	G	U	A	No	–	–
[93]	Newspapers	B	U	R	No	–	78–100
[95]	Journals, forms	B	U	A	Yes	–	–
[109]	Book and CD covers, video	C	C	A	No	–	–
[110]	WWW images	C	C	A	No	262	90

Table 2: Properties related to the processing time for document analysis methods.

Ref.	Document type	Time	Image size	Res.	Platform
[3]	Journals, newspapers	80	A4	400	–
[7]	Journals, newspapers, newsletters, advertisements	0.55	810 × 1151	100	HP 9000/735
[7]	Journals, newspapers, newsletters, advertisements	1.2	1215 × 1727	150	HP 9000/735
[7]	Journals, newspapers, newsletters, advertisements	6.5	2431 × 3455	300	HP 9000/735
[17]	Journal covers	95	2000 × 2679	–	Pentium 100
[17]	Journal covers	180	1719 × 2476	–	Pentium 100
[26]	Journals, business cards, technical reports	4.8	–	–	PC 486
[30]	Journals	22	A4	300	Sun Sparc 20
[34]	Newspapers	≈9	6592 × 9890	–	Pentium 350 II
[35]	Journals	≈2	Letter-sized	300	Sun Sparc 10
[42]	Advertisements, envelopes, letters	306.9-563.6	<A4	300	Sun Sparc IPX
[45]	Newspapers	≈120	512 × 512	75	Sun Sparc 2
[46]	Journals	1.3	A4	300	SG Indigo
[48]	Advertisements	0.15	548 × 769	150	Sun Ultra-Sparc I
[48]	WWW images	0.11	385 × 234	–	Sun Ultra-Sparc I
[48]	Book covers	0.4	763 × 537	50	Sun Ultra-Sparc I
[48]	Video frames	0.09	160 × 120	–	Sun Ultra-Sparc I

Table 2 continued

Ref.	Document type	Time	Image size	Res.	Platform
[49]	Journals	60-85	264 × 332 to 780 × 1080	100	Sun Sparc 20
[53]	Journals, newspapers	2.93	1053 × 1149	90	Pentium 200 Pro
[53]	Journals, newspapers	5.37-7.03	2592 × 3300– 3114 × 3554	300	Pentium 200 Pro
[68]	Book covers	0.01-2.86	512 × 512	–	Pentium 200 Pro
[70]	Journals	20	1278 × 1746	–	Sun Sparc 2
[74]	Journals	0.9-1.9	A4	300	Sun Sparc
[80]	Journals, newspapers	2.37	A4	300	Sun Sparc
[84]	Book and journal covers	21.31	1600 × 2400	200	SunUltra Sparc 5/10
[93]	Newspapers	2.6	–	100	Sun 3/60
[93]	Newspapers	9.5	–	200	Sun 3/60
[109]	Book and CD covers, video	5.5-6	256 × 256	–	Sun Sparc 20

5 Analysis of the methods

In this section we try to generalize our ideas about document layout analysis techniques. Although some authors [42, 44, 48, 62, 68, 95, 109] state that their methods are applicable to a variety of document classes and image types, it seems that there is no generic solution to the document segmentation and classification problem, because the broader a task is, the less manageable it is, the more parameters have to be adjusted, and the less predictable the results.

This can be seen especially for the methods developed for text extraction from complex gray-scale or color images. Usually such methods rely much more on heuristics than those used for other classes. A common feature of these methods is that a number of conditions must be satisfied, such as 1) uniformity of character color, 2) well-separated colors of text and background, 3) the characters should form a straight horizontal line (the last condition becomes unnecessary for the recently proposed methods [37, 58, 68, 73]).

Text extraction from color images does not use OCR to verify the extraction results. Application of OCR to this task might sometimes reduce the number of heuristics used and improve the accuracy.

The layout analysis methods applied to images of journals, newspapers and other text-dominated documents can be divided into two groups. The methods in the first group require skew estimation and correction before layout analysis. These methods are typically applied to images with a rectangular or Manhattan layout, because skew correction for such images significantly facilitates layout analysis. However, errors in skew detection will degrade the accuracy of a layout analysis method if it cannot operate on skewed regions. On the other hand, another group of methods first segment and classify the image into regions and then estimate and correct the skew of text regions. Usually such methods work with complex and arbitrary layouts or with multiple skew. In the latter case, skew estimation applied to all text blocks may not be useful. Sometimes these methods avoid extra processing, because skew estimation is not done for non-text regions. However, layout analysis seems to be a more difficult task than skew estimation, and it is not always easy to do it as accurately as skew estimation. This means that errors in complex layout analysis may appear more frequently than those in skew estimation, resulting in postprocessing after skew estimation. The final choice depends on the application and the complexity of the problem.

The initial image resolution is different in different applications. For example, it is low (72 dpi) for WWW images [110], while it can be quite high (up to 300 dpi) for journal or newspaper images. Processing at low resolution results in fast computation, but fine document details are corrupted and this may (though does not always) negatively influence other steps such as text segmentation or OCR. The choice of a proper resolution is therefore not a simple task because it involves analysis of many of the document processing steps.

Data structures are of great importance for layout analysis methods. They can dramatically reduce the processing time and determine how easily the data can be accessed in the operations after layout analysis. Examples of flexible and efficient structures are the BAG [46, 47], white tile [5–7], quadtree [62], and square block tessellation [16, 29, 30, 78]. A good data representation not only provides not only easy access to the data, but often results in skew- and/or layout-independence.

There have been only a few applications of soft computing techniques to document layout analysis [29, 30, 49, 56]. The paper [56] describes comprehensive research using several popular

neural networks for document classification. However, the number of training samples can be very large (up to 1,000,000 in [49]) and it is unclear how many samples are needed because of large inter-class and intra-class variations.

The papers [16, 49, 62] demonstrate the effectiveness of document layout analysis methods when solving very difficult tasks of text extraction from a complex textured background [16, 62] and separation of different languages within the same image [49]. Researchers in document image analysis have not yet paid much attention to these tasks.

Use of the background can greatly help segmentation, because it is a natural separator between different regions. The image resolution can often be reduced to 75–100 dpi before processing. It is also better to combine document segmentation with document classification; this saves much processing time because the data need to be accessed only once.

Processing time and accuracy are important features of document layout analysis methods. Processing time varies widely: from $\approx$1 s to several minutes per image. We have found that fast methods such as [7, 53] which are simultaneously skew- and layout-independent take approximately 5.5–7 s to process a binary image at 300 dpi. Faster methods [46, 47, 74] take 1–2 s to do this, but in this case, either prior skew correction is necessary or document layout cannot be arbitrary. Fast computation may be more important for information retrieval, such as text extraction from WWW images or video, than for the analysis of journal and newspaper images, which can be more or less interactive. In the latter case, the user often has more freedom to edit the results without new parameter settings than in the former case, where processing with new settings is necessary if the previous results are not satisfactory. Thus how fine parameter settings should be depends on the particular application.

It is quite difficult to compare the accuracies of different methods, because often they are tested on different data sets with different initial conditions. In many cases, except for the analysis of pure text images, the results are only visually evaluated and objective performance evaluation is often absent. Moreover, there is no unique definition of an accuracy measure. For example, it can be the ratio of the number of regions/pages correctly segmented and classified to the total number of regions/pages, or it can be expressed by the number of cases where regions are erroneously split or merged. Splitting of a region containing data of the same class into several subregions may often be more easily corrected at later steps than merging of regions belonging to different classes, which is not easy to detect. The accuracy varies from 70 to almost 100 percent for various methods and various document classes. It is higher for binary images

of journals and newspapers and lower for color WWW and video images.

Ground-truthing and benchmarking of document layout analysis methods is not completely solved. Many methods are claimed to be skew-invariant, but it is very difficult to verify their performance on a large data set of skewed images, because ground truth is usually created only for upright images without skew. Currently available benchmarking systems evaluate performance based on OCR results [51] (here it is not clear whether a mistake is due to bad segmentation or to incorrect character recognition), or they need separate ground truth for each skewed image [8, 52, 102, 103]. In the latter case, an image has to be scanned at all possible skew angles, and after that ground-truth is generated for each of them. Automated and accurate ground-truthing is very time-consuming (up to 5-10 min per image at the pixel level as reported in [102, 103], and up to 5 min per image at the bounding box level as reported in [52]). Therefore, if one needs to evaluate the performance of a method using only one image skewed at all angles from 1 to 100 degrees in 1 degree, steps, it is necessary to scan this image 100 times at different angles. In this case, ground-truthing could take many hours! That is why many authors prefer the old method of evaluation—their own visual perception. However, promising results on evaluation of layout analysis algorithms and ground truth generation have begun to appear [8, 50, 52]. The method in [52] can be especially useful for ground truth generation because it allows one to do it automatically, but it seems that it can be primarily applied to text documents without pictures or graphics.

In practice, it would be better to use specialized algorithms for different types of documents and tasks in order to get optimal performance. The following features would be useful in any case:

- tolerance to skew (uniform and non-uniform),
- layout independence,
- text extraction both on white and on inverse backgrounds,
- easy access to data,
- fast speed and high accuracy,
- independence of font type/size and script.

6 Conclusions

This survey has overviewed the state of the art in document layout analysis and has described the progress in this area primarily in the 1990's. First, we presented a brief analysis of the tasks needed for various document classes. Then we focused in more detail on three groups: 1) structured articles, 2) documents with unconstrained layout, and 3) non-traditional documents. In addition to describing the methods used for these groups, we considered such features as image and background type, image resolution, processing time, tolerance to skew, and different layouts.

Despite intensive research in this area, there is still no general method of processing the images of different document classes both accurately and automatically. Important features that such a method can have are skew-, layout-, and script-independence, fast speed, high accuracy, and flexible image representation enabling easy access to the data. Proper formalization of notions of "graphic" and "picture", automatic ground-truth generation, benchmarking of layout analysis methods for binary/gray-scale images, and more automated and less heuristic-based color image processing ought to be among the goals of future research.

References

[1] A. Agarwal, L. Granowetter, K. Hussein, and A. Gupta, Detection of courtesy amount block on bank checks, in: *Proceedings of the 3rd ICDAR* (Montréal, Canada, 1995) 748–751.

[2] C. Ah-Soon and K. Tombre, Variations on the analysis of architectural drawings, in: *Proceedings of the 4th ICDAR* (Ulm, Germany, 1997) 347–351.

[3] T. Akiyama and N. Hagita, Automated entry system for printed documents, *Pattern Recognition* **23** (1990) 1141–1154.

[4] M. Anegawa, O. Shiku, A. Nakamura, T. Ohyama, and H. Kuroda, A system for recognizing numeric strings from topographical maps, in: *Proceedings of the 3rd ICDAR* (Montréal, Canada, 1995) 940–943.

[5] A. Antonacopoulos and R.T. Ritchings, Flexible page segmentation using the background, in: *Proc. of the 12th ICPR* (Jerusalem, Israel, 1994) 339–344.

[6] A. Antonacopoulos and R.T. Ritchings, Representation and classification of complex-shaped printed regions using white tiles, in: *Proceedings of the 3rd ICDAR* (Montréal, Canada, 1995) 1132–1135.

[7] A. Antonacopoulos, Page segmentation using the description of the background, *Computer Vision and Image Understanding* **70** (1998) 350–369.

[8] A. Antonacopoulos and A. Brough, Methodology for flexible analysis of the performance of page segmentation algorithms, in: *Proceedings of the 5th ICDAR* (Bangalore, India, 1999) 451–454.

[9] H. Arai and K. Odaka, Form processing based on background region analysis, in: *Proceedings of the 4th ICDAR* (Ulm, Germany, 1997) 164–169.

[10] J.F. Arias, R. Kasturi, and A. Chhabra, Efficient techniques for telephone company line drawing interpretation, in: *Proceedings of the 3rd ICDAR* (Montréal, Canada, 1995) 795–798.

[11] Y. Ariki, K. Matsuura, and S. Takao, Telop and flip frame detection and character extraction from TV news articles, in: *Proceedings of the 5th ICDAR* (Bangalore, India, 1999) 701–704.

[12] H.S. Baird, Background structure in document images, in: *Advances in Structural and Syntactic Pattern Recognition* (World Scientific, Singapore, 1992) 253–269.

[13] T.A. Bayer and H.U. Mogg-Schneider, A generic system for processing invoices, in: *Proceedings of the 4th ICDAR* (Ulm, Germany, 1997) 740–744.

[14] F. Cesarini, M. Gori, G. Soda, and S. Marinai, A system for data extraction from forms of known class, in: *Proceedings of the 3rd ICDAR* (Montréal, Canada, 1995) 1136–1140.

[15] F. Cesarini, M. Gori, S. Marinai, and G. Soda, Structured document segmentation and representation by the modified X-Y tree, in: *Proceedings of the 5th ICDAR* (Bangalore, India, 1999) 563–566.

[16] J.-L. Chen, A simplified approach to the HMM based texture analysis and its application to document segmentation, *Pattern Recognition Letters* **18** (1997) 993–1007.

[17] W.-Y. Chen and S.-Y. Chen, Adaptive page segmentation for color technical journals' cover images, *Image and Vision Computing* **16** (1998) 855–877.

[18] M. Cheriet, J.N. Said, and C.Y. Suen, A formal model for document processing of business forms, in: *Proceedings of the 3rd ICDAR* (Montréal, Canada, 1995) 210–213.

[19] Y.-H. Chiou and H.-J. Lee, Recognition of Chinese business cards, in: *Proceedings of the 4th ICDAR* (Ulm, Germany, 1997) 1028–1032.

[20] L. Cinque, L. Lombardi, and G. Manzini, A multiresolution approach for page segmentation, *Pattern Recognition Letters* **19** (1998) 217–225.

[21] C. Cracknell, A.C. Downton, and L. Du, An object-oriented form description language and approach to handwritten form processing, in: *Proceedings of the 4th ICDAR* (Ulm, Germany, 1997) 180–184.

[22] M. Cullen, L. Pintsov, and B. Romansky, Reading encrypted postal indicia, in: *Proceedings of the 3rd ICDAR* (Montréal, Canada, 1995) 1018–1023.

[23] A.K. Das and N.A. Langrana, Recognition of dimension sets and integration with vectorized engineering drawings, in: *Proceedings of the 3rd ICDAR* (Montréal, Canada, 1995) 347–350.

[24] O. Déforges and D. Barba, Segmentation of complex documents multilevel images: a robust and fast text bodies-headers detection and extraction scheme, in: *Proceedings of the 3rd ICDAR* (Montréal, Canada, 1995) 770–773.

[25] S. Djeziri, F. Nouboud, and R. Plamondon, Extraction of items from checks, in: *Proceedings of the 4th ICDAR* (Ulm, Germany, 1997) 749–752.

[26] D. Drivas and A. Amin, Page segmentation and classification utilising bottom-up approach, in: *Proceedings of the 3rd ICDAR* (Montréal, Canada, 1995) 610–614.

[27] V. Edlin and H. Emptoz, Logarithmic spiral grid and gaze control for the development of strategies of visual segmentation on a document, in: *Proceedings of the 4th ICDAR* (Ulm, Germany, 1997) 689–692.

[28] L. Eikvil, K. Aas, and H. Koren, Tools for interactive map conversion and vectorization, in: *Proceedings of the 3rd ICDAR* (Montréal, Canada, 1995) 927–930.

[29] K. Etemad, D. Doermann, and R. Chellappa, Page segmentation using decision integration and wavelet packets, in: *Proceedings of the 12th ICPR* (Jerusalem, Israel, 1994) 345–349.

[30] K. Etemad, D. Doermann, and R. Chellappa, Multiscale segmentation of unstructured document pages using soft decision integration, *IEEE Trans. on PAMI* **19** (1997) 92–96.

[31] K.-C. Fan, C.-H. Liu, and Y.-K. Wang, Segmentation and classification of mixed text/graphics/image documents, *Pattern Recognition Letters* **15** (1994) 1201–1209.

[32] K.-C. Fan and L.-S. Wang, Classification of document blocks using density feature and connectivity histogram, *Pattern Recognition Letters* **16** (1995) 955–962.

[33] U. Gargi, D. Crandall, A. Antani, T. Gandhi, R. Keener, and R. Kasturi, A system for automatic text detection in video, in: *Proceedings of the 5th ICDAR* (Bangalore, India, 1999) 29–32.

[34] B. Gatos, S.L. Mantzaris, K.V. Chandrinos, A. Tsigris, and S.J. Perantonis, Integrated algorithms for newspaper page decomposition and article tracking, in: *Proceedings of the 5th ICDAR* (Bangalore, India, 1999) 559–562.

[35] J. Ha, R.M. Haralick, and I.T. Phillips, Recursive X-Y cut using bounding boxes of connected components, in: *Proceedings of the 3rd ICDAR* (Montréal, Canada, 1995) 952–955.

[36] R.M. Haralick, Document image understanding: geometrical and logical layout, in: *Proceedings of CVPR'94* (Seattle, WA, 1994) 385–390.

[37] H. Hase, T. Shinokawa, M. Yoneda, M. Sakai, and H. Maruyama, Character string extraction from a color document, in: *Proceedings of the 5th ICDAR* (Bangalore, India, 1999) 75–78.

[38] Y. Hirayama, A method for table structure analysis using DR matching, in: *Proceedings of the 3rd ICDAR* (Montréal, Canada, 1995) 583–586.

[39] O. Hori, A video text extraction method for character recognition, in: *Proceedings of the 5th ICDAR* (Bangalore, India, 1999) 25–28.

[40] G.F. Houle, D.B. Aragon, R.W. Smith, M. Shridhar, and D. Kimura, A multi-layered corroboration-based check reader, in: *Proceedings of DAS'96* (Malvern, PA, 1996) 495–546.

[41] G. Hutton, M. Cripps, D.G. Elliman, and C.A. Higgins, A strategy for on-line interpretation of sketched engineering drawings, in: *Proceedings of the 4th ICDAR* (Ulm, Germany, 1997) 771–775.

[42] F. Hönes and J. Lichter, Layout extraction of mixed mode documents, *Machine Vision and Applications* **7** (1994) 237–246.

[43] Y. Ishitani, Model matching based on association graph for form image understanding, in: *Proceedings of the 3rd ICDAR* (Montréal, Canada, 1995) 287–292.

[44] Y. Ishitani, Document layout analysis based on emergent computation, in: *Proceedings of the 4th ICDAR* (Ulm, Germany, 1997) 45–50.

[45] A.K. Jain and S. Bhattacharjee, Text segmentation using Gabor filters for automatic document processing, *Machine Vision and Applications* **5** (1992) 169–184.

[46] A.K. Jain and B. Yu, Page segmentation using document model, in: *Proceedings of the 4th ICDAR* (Ulm, Germany, 1997) 34–38.

[47] A.K. Jain and B. Yu, Document representation and its application to page decomposition, *IEEE Trans. on PAMI* **20** (1998) 294–308.

[48] A.K. Jain and B. Yu, Automatic text location in images and video frames, *Pattern Recognition* **31** (1998) 2055–2076.

[49] A.K. Jain and Y. Zhong, Page segmentation using texture analysis, *Pattern Recognition* **29** (1996) 743–770.

[50] M. Junker, R. Hoch, and A. Dengel, On the evaluation of document analysis components by recall precision and accuracy, in: *Proceedings of the 5th ICDAR* (Bangalore, India, 1999) 713–716.

[51] J. Kanai, S.V. Rice, T.A. Nartker, and G. Nagy, Automatic evaluation of OCR zoning, *IEEE Trans. on PAMI* **17** (1995) 86–90.

[52] T. Kanungo and R.M. Haralick, An automatic closed-loop methodology for generating character groundtruth for scanned documents, *IEEE Trans. on PAMI* **21** (1999) 179–183.

[53] K. Kise, A. Sato, and M. Iwata, Segmentation of page images using the area Voronoi diagram, *Computer Vision and Image Understanding* **70** (1998) 370–382.

[54] S. Knerr, E. Augustin, O. Buret, and D. Price, Hidden Markov Model based word recognition and its application to legal amount reading on French checks, *Computer Vision and Image Understanding* **70** (1998) 404–419.

[55] M. Köppen, D. Waldöstl, and B. Nickolay, A system for the automated evaluation of invoices, in: *Proceedings of DAS'96* (Malvern, PA, 1996) 3–21.

[56] D.X. Le, G.R. Thoma, and H. Wechsler, Classification of binary document images into textual or non-textual data blocks using neural network models, *Machine Vision and Applications* **8** (1995) 289–304.

[57] E. Lethelier, M. Leroux, and M. Gilloux, An automatic reading system for handwritten numeral amounts on French checks, in: *Proceedings of the 3rd ICDAR* (Montréal, Canada, 1995) 92–97.

[58] H. Li, D. Doermann, and O. Kia, Automatic text detection and tracking in digital video, to appear in *IEEE Trans. on Image Processing - Special Issue on Image and Video Processing for Digital Libraries* (1999).

[59] H. Li, O. Kia, and D. Doermann, Text enhancement in digital video, in: *Proceedings of SPIE Conf. on Document Recognition and Retrieval VI* (San Jose, CA, 1999) 2–9.

[60] L. Li, G. Nagy, A. Samal, S. Seth, and Y. Xu, Cooperative text and line-art extraction from a topographic map, in: *Proceedings of the 5th ICDAR* (Bangalore, India, 1999) 467–470.

[61] J. Liu, X. Ding, and Y. Wu, Description and recognition of form and automated form data entry, in: *Proceedings of the 3rd ICDAR* (Montréal, Canada, 1995) 579–582.

[62] J. Liu, Y.Y. Tang, and C.Y. Suen, Chinese document layout analysis based on adaptive split-and-merge and qualitative spatial reasoning, *Pattern Recognition* **30** (1997) 1265–1278.

[63] D. Lopresti and J. Zhou, Document analysis and the world wide web, in: *Proceedings of DAS'96* (Malvern, PA, 1996) 651–671.

[64] W. Lu, W. Wu, and M. Sakauchi, A drawing recognition system with rule acquisition ability, in: *Proceedings of the 3rd ICDAR* (Montréal, Canada, 1995) 512–515.

[65] H. Luo, G. Agam, and I. Dinstein, Directional mathematical morphology approach for line thinning and extraction of character strings from maps and line drawings, in: *Proceedings of the 3rd ICDAR* (Montréal, Canada, 1995) 257–260.

[66] H. Luo, R. Kasturi, J.F. Arias, and A. Chhabra, Interpretation of lines in distributing frame drawings, in: *Proceedings of the 4th ICDAR* (Ulm, Germany, 1997) 66–70.

[67] S. Madhvanath, V. Govindaraju, V. Ramanaprasad, D.S. Lee, and S.N. Srihari, Reading handwritten US census forms, in: *Proceedings of the 3rd ICDAR* (Montréal, Canada, 1995) 82–85.

[68] S. Messelodi and C.M. Modena, Automatic identification and skew estimation of text lines in real scene images, *Pattern Recognition* **32** (1999) 791–810.

[69] U. Miletzki, Documents on the move— DA&IR-driven mail piece processing today and tomorrow, in: *Proceedings of DAS'96* (Malvern, PA, 1996) 547–563.

[70] N. Normand and C. Viard-Gaudin, A background based adaptive page segmentation algorithm, in: *Proceedings of the 3rd ICDAR* (Montréal, Canada, 1995) 138–141.

[71] J.M. Ogier, R. Mullot, J. Labishe, and Y. Lecourtier, Multilevel approach and distributed consistency for technical map interpretation: application to cadastral maps, *Computer Vision and Image Understanding* **70** (1998) 438–451.

[72] O. Okun and M. Pietikäinen, A survey of texture-based methods for document layout analysis, in: *Proceedings of Workshop on Texture Analysis in Machine Vision* (Oulu, Finland, 1999) 137–148.

[73] T. Park, D. Kim, and K. Chung, Orientation and scale invariant text region extraction in WWW images, in: *Proceedings of IAPR Workshop on Machine Vision and Applications* (Chiba, Japan, 1998) 290–293.

[74] T. Pavlidis and H. Zhou, Page segmentation and classification, *Computer Vision, Graphics, and Image Processing* **54** (1992) 484–496.

[75] J.S. Payne, T.J. Stonham, and D. Patel, Document segmentation using texture analysis, in: *Proceedings of the 12th ICPR* (Jerusalem, Israel, 1994) 380–382.

[76] M. Pierrot-Deseilligny, H. Le Men, and G. Stamon, Characters string recognition on maps: a method for high level reconstruction, in: *Proceedings of the 3rd ICDAR* (Montréal, Canada, 1995) 249–252.

[77] R. Safari, N. Narasimhamurthi, M. Shridhar, and M. Ahmadi, Form registration: a computer vision approach, in: *Proceedings of the 4th ICDAR* (Ulm, Germany, 1997) 758–761.

[78] J. Sauvola and M. Pietikäinen, Page segmentation and classification using fast feature extraction and connectivity analysis, in: *Proceedings of the 3rd ICDAR* (Montréal, Canada, 1995) 1127–1131.

[79] J.J. Shieh, Recursive morphological sieve method for searching pictorial point symbols on maps, in: *Proceedings of the 3rd ICDAR* (Montréal, Canada, 1995) 931–935.

[80] F.Y. Shih and S.-S. Chen, Adaptive document block segmentation and classification, *IEEE Trans. on SMC - Part B: Cybernetics* **26** (1996) 797–802.

[81] M. Shridhar, J.W.V. Miller, G. Houle, and L. Bijnagte, Recognition of license plate images: issues and perspectives, in: *Proceedings of the 5th ICDAR* (Bangalore, India, 1999) 17–20.

[82] L. Simoncini and Zs.M. Kovács, A system for reading USA census'90 hand-written fields, in: *Proceedings of the 3rd ICDAR* (Montréal, Canada, 1995) 86–91.

[83] R. Sivaramakrishnan, I.T. Phillips, J. Ha, S. Subramanium, and R.M. Haralick, Zone classification in a document using the method of feature vector generation, in: *Proceedings of the 3rd ICDAR* (Montréal, Canada, 1995) 541–544.

[84] K. Sobottka, H. Bunke, amd H. Kronenberg, Identification of text on colored book and journal covers, in: *Proceedings of the 5th ICDAR* (Bangalore, India, 1999) 57–62.

[85] W. Song, M. Feng, and X. Shaowei, A Chinese bank check recognition system based on the fault tolerant technique, in: *Proceedings of the 4th ICDAR* (Ulm, Germany, 1997) 1038–1042.

[86] S.N. Srihari, T. Hong, and G. Srikantan, Machine printed Japanese document recognition, *Pattern Recognition* **30** (1997) 1301–1313.

[87] C.L. Tan and P.O. Ng, Text extraction using pyramid, *Pattern Recognition* **31** (1998) 63–72.

[88] Y.Y. Tang, H. Ma, X. Mao, D. Liu, and C.Y. Suen, A new approach to document analysis based on modified fractal signature, in: *Proceedings of the 3rd ICDAR* (Montréal, Canada, 1995) 567–570.

[89] Y.Y. Tang, S.-W. Lee, and C.Y. Suen, Automatic document processing: a survey, *Pattern Recognition* **29** (1996) 1931–1952.

[90] L.Y. Tseng and R.-C. Chen, The recognition of form documents based on three types of line segments, in: *Proceedings of the 4th ICDAR* (Ulm, Germany, 1997) 71–75.

[91] Ø.D. Trier, T. Taxt, and A.K. Jain, Data capture from maps based on gray scale topographic analysis, in: *Proceedings of the 3rd ICDAR* (Montréal, Canada, 1995) 923–926.

[92] H. Walischewski, Learning regions of interest in postal automation, in: *Proceedings of the 5th ICDAR* (Bangalore, India, 1999) 317–320.

[93] D. Wang and S.N. Srihari, Classification of newspaper image blocks using texture analysis, *Computer Vision, Graphics, and Image Processing* **47** (1989) 327–352.

[94] J.-Y. Wang, L.-H. Chen, K.-C. Fan, and H.-Y.M. Liao, Separation of Chinese characters from graphics, in: *Proceedings of the 3rd ICDAR* (Montréal, Canada, 1995) 948–951.

[95] S.Y. Wang and T. Yagasaki, Block selection: a method for segmenting page image of various editing styles, in: *Proceedings of the 3rd ICDAR* (Montréal, Canada, 1995) 128–133.

[96] X. Wang and T. Tsutsumida, A new method of character line extraction from mixed-unformatted document image for Japanese mail address recognition, in: *Proceedings of the 5th ICDAR* (Bangalore, India, 1999) 769–772.

[97] T. Watanabe and X. Huang, Automatic acquisition of layout knowledge for understanding business card, in: *Proceedings of the 4th ICDAR* (Ulm, Germany, 1997) 216–220.

[98] L. Wenyin and D. Dori, Automated CAD conversion with the machine drawing understanding system, in: *Proceedings of DAS'96* (Malvern, PA, 1996) 241–259.

[99] M. Wolf, H. Niemann, and W. Schmidt, Fast address block location on handwritten and machine printed mail-piece images, in: *Proceedings of the 4th ICDAR* (Ulm, Germany, 1997) 753–757.

[100] L. Xingyuan, D. Doermann, W.-G. Oh, and W. Gao, A robust method for unknown forms analysis, in: *Proceedings of the 5th ICDAR* (Bangalore, India, 1999) 531–534.

[101] J. Xue, X. Ding, C. Liu, S. Pan, and H. Kong, Destination address block location on handwritten Chinese envelope, in: *Proceedings of the 5th ICDAR* (Bangalore, India, 1999) 737–740.

[102] B.A. Yanikoglu and L. Vincent, Ground-truthing and benchmarking document page segmentation, in: *Proceedings of the 3rd ICDAR* (Montréal, Canada, 1995) 601–604.

[103] B.A. Yanikoglu and L. Vincent, Pink Panther: a complete environment for ground-truthing and benchmarking document page segmentation, *Pattern Recognition* **31** (1998) 1191–1204.

[104] B. Yu, A.K. Jain, and M. Mohiuddin, Address block location on complex mail pieces, in: *Proceedings of the 4th ICDAR* (Ulm, Germany, 1997) 897–901.

[105] C.L. Yu, C.Y. Suen, and Y.Y. Tang, Location and recognition of legal amounts on Chinese bank cheques, in: *Proceedings of the 4th ICDAR* (Ulm, Germany, 1997) 588–591.

[106] Y. Yu, A. Samal, and S.C. Seth, A system for recognizing a large class of engineering drawings, in: *Proceedings of the 3rd ICDAR* (Montréal, Canada, 1995) 791–794.

[107] J. Yuan, Y.Y. Tang, and C.Y. Suen, Four directional adjacency graphs (FDAG) and their application in locating fields in forms, in: *Proceedings of the 3rd ICDAR* (Montréal, Canada, 1995) 752–755.

[108] Y. Zhong, K. Karu, and A.K. Jain, Locating text in complex color images, in: *Proceedings of the 3rd ICDAR* (Montréal, Canada, 1995) 146–149.

[109] Y. Zhong, K. Karu, and A.K. Jain, Locating text in complex color images, *Pattern Recognition* **28** (1995) 1523–1535.

[110] J. Zhou and D. Lopresti, Extracting text from WWW images, in: *Proceedings of the 4th ICDAR* (Ulm, Germany, 1997) 248–252.