Document Image Understanding: Geometric and Logical Layout

Robert M. Haralick
Electrical Engineering FT-10
University of Washington
Seattle, WA 98115

1 Introduction

Document Image Understanding encompasses the technology required to make paper documents equivalent to other computer exchange media like floppies, tapes, and cdroms. The physical reader of the paper document is the scanner just like the physical reader of the floppy is the floppy drive and the physical reader of the tape cartridge is the tape cartridge drive, and the physical reader of the cdrom is the cdrom drive.

But document image understanding can involve more than just reading the character strings on a paper document and putting them in a format of our favorite word processing system. For documents have on them information just like floppies have information. But the information on a floppy is relatively simple. Its structure is typically just a set of files. But paper documents have a much more complicated structure. For example a business letter has a sender's address, a receiver's address, a date, an opening salutation, a body, a closing, and a signature. A technical document has a title, one or more authors, an abstract, key words, technical sections, displayed equations, tables, graphs, illustrations, footnotes, page numbers, a reference list, the conference or journal or institution of publication, and a date. And the purpose of inputting a technical document or letter into a computer is not just to transform it into computer readable form, but to put it in a form by which its information can be queried. Hence determining what kind of document it is and determining its logical structure: what areas are non-text or text, determining the read order of the text blocks, and determining information like title, author, etc. constitute important aspects of document image understanding.

Understanding paper documents is further complicated by the fact that paper documents not only hold information which is put in character strings. But the information can be in a drawing or in graphics. So understanding line drawings, engineering line drawings, perspective projection drawings, graphs, and special kinds of document like music scores are all part of document image understanding.

In this survey we restrict ourselves to documents such as business letters, forms, and scientific and technical articles such as those found in archival journals and technical conferences. Understanding such documents involves estimating the rotation skew of each document page, determining the geometric page layout, labeling blocks as text or non-text, determining the read order for text blocks, recognizing the text of text blocks through an OCR system, determining the logical page layout, and formatting the data and information of the document in a suitable way for use by a word processing system or by an information retrieval system.

Document image understanding work not covered in this survey include machine printed character recognition, hand printed character recognition, cursive script recognition, musical score recognition, map interpretation, engineering drawing interpretation, online character recognition, script determination, language determination, contextual character recognition, word recognition, graphic drawing interpretation, postal address location, and table understanding.

Papers covering all areas of document image can be found in the 1st and 2nd International Conference on Document Analysis and Recognition [ICDAR] (1991 and 1993), the 1992, 1993, and 1994 Annual Symposiums on Document Analysis and Infomation Retireval sponsored by the University of Nevada, Las Vegas, and the International Conferences on Pattern Recognition [ICPR]. The brevity of the space for this survey necessitates that many papers from these conferences are not mentioned. Surveys of document image analysis and document image understanding can be found in (Casey and Nagy, 1991; and Tang et. al. 1991).

2 Geometric Page Layout

A geometric page layout of a document image page is a specification of the geometry of the maximal homogeneous regions and the spatial relations of these regions. A region is homogeneous if all its area is of one type: text, or figure etc. and each text line of the page lies entirely within some text region of the layout. Formally, a *geometric page layout* $\Phi = (\mathcal{R}, \mathcal{S})$, where $\mathcal{R}$ is a set of regions and $\mathcal{S}$ is a labeled spatial relation on the set $\mathcal{R}$ of regions. A *region* $R = (T, \theta)$, where T defines the type of the region and θ is the parameter vector of values for the region (see Table 1). The parameter vector can as well include uncertainty for any of the parameters. Uncertainty can be a parameter standard deviation, or a tolerance interval - probability pair, or for the entire parameter vector the uncertainty can be specified by a covariance matrix.

The labeled spatial relation $\mathcal{S} \subseteq \mathcal{R} \times \mathcal{R} \times \mathcal{L}$ contains triples (R_1, R_2, L), where R_1 is one region, R_2 is a second region, and L is a label. The triple means

T	θ
circle	radius, position of center
ellipse	major and minor axis length, major axis orientation, center position
square	side length, orientation of one side, center position
rectangle	two side lengths, orientation of first side, center position
polygon	number of sides, side lengths, orientation of first side, center position

Table 1: Typical Values For Region Types and Associated Parameter Vectors

that region R_1 stands in relation L with respect to R_2. L can be atomic, taking values like *inside of, mutually exclusive of,* or *overlapping with.* Or L can be complex taking values like *the difference between the X-coordinate of the center of side i of region* R_1 *and the X-coordinate of the center of side j of region* R_2 *is between* α_1 *and* α_2.

A Manhattan page layout is one where the regions of the page layout are all rectangular and the rectangles are in the same orientation. Hence after an appropriate page rotation the sides of the rectangles will all be either horizontal or vertical. Furthermore, each pair of rectangles either is mutually exclusive or one contains the other.

Fujisawa and Nakano (1990), Higashino et. al. (1986), and Yu et. al. (1993) discuss page layout representations which are specializations of the formalism presented here.

The purpose of a text-block segmentation is to determine a page layout whose regions cover the page such that each region, excluding its holes, (the regions which it contains), is entirely a text region or a non-text region and such that in each text region no text line is fragmented. Regions which have been identified as text regions can then be given to an OCR system to produce the computer readable text within the region.

Many of the algorithms for determining geometric layout employ the operations of mathematical morphology. Although the original algorithm developers generally do not describe their algorithms in terms of mathematical morpholgy and were probably not aware that their algorithm han anything to do with mathematical morphology, our description will utilize the operations of mathematical. This will allow us to be brief and precise.

Early work on page segmentation was done by Wahl et. al. (1982). They use a technique called the constrained run length smoothing algorithm. It actually consists of a morphologically closing of the document image with a horizontal structuring element of specified length (they used 300) intersected with a morphological closing of the document image with a vertical structuring element of specified length (they used 500). The intersection is then morphologically closed with a horizontal structuring element of specified length (they used 30). The bounding rectangles of the connected components of the resulting image constitute the block segments. Features of the blocks include the area of the connected component of the block, the number of black pixels in the block on the original document image. the mean horizontal black run lengths of the original image within the blocks, and the height and width of the bounding rectangle of the block. Text areas are classified into *text, horizontal solid black lines, graphic and halftone images,* and *vertical solid black lines.* No measure of performance is given.

Nagy and Seth (1984), and Nagy et. al. (1986) employ an $X-Y$ tree as the representation of a page layout. The root node of an $X-Y$ tree is the bounding rectangle of the full page. Each node in the tree represents a rectangle in the page. The children of a node are obtained by subdividing the rectangle of the parent node either horizontally or vertically, with horizontal and vertical cuts being alternately employed in successive levels in the tree. Hao et. al. (1993) describe a variation on this technique.

Fisher et. al. (1990) sample a 300dpi document image by a factor of 4 and use the run length smoothing algorithm. They then compute the connected components of the run length smoothed image. The connected components and their bounding boxes constitute the blocks of the geometric page layout. They extract connected component features such as component height, width, aspect ratio, density, perimeter, and area for classifying each block as *text* or *non-text.*

Lebourgeois et.al. (1992) sample the document image by a factor of 8 vertically and 3 horizontally. Each pixel on the sampled image corresponds to an 8x3 window on the original image. If any pixel on the 8x3 window of the original image is a binary one then the sampled image has a binary one in the corresponding pixel position. Then the sampled image is dilated by a horizontal structuring element to effectively smear adjacent characters into one another. Each connected component is then characterized by its bounding rectangle and the mean horizontal length of the black runs. Connected components having a vertical height within given bounds and mean horizontal run length within given bounds are then labeled as a text lines and outside the given bounds are labeled as a non-text lines. Components assigned as text regions are then vertically merged into larger blocks using rules taking into account alignment. Blocks are also subdivided to separate their horizontal peninsulas. No measure of performance is given but an indication that the method needs improvement was stated.

Bloomberg (1991) uses morphological operations on the document image at various resolutions to determine identify font style for each word. Class labels include *bold, italic,* and *normal.* The method employs a small vertical dilation followed by a close open sequence to remove noise followed by a hit and miss transform to identify seed points of characters in the italic class or bold class. Then the words which are in italic or bold can be delineated by conditionally dilating the seed with a precalculated word segmentation mask. No accuracy performance results are given.

Saitoh and Pavlidis (1992) proceed sampling by 8 vertically and 4 horizontally and then extracting the connected components. They then classify each component into *text, text or noise, diagram or table, halftone image, horizontal separator*, or *vertical separator*, using block attributes such as block height, height to width ratio, and connectivity features of the line adjacency graph, and whether there are vertical or horizontal rulings. Page rotation skew is estimated from a least squares line fit to the the center points of blobs belonging to the same block. Blocks are subdivided based on the vertical distance between lines in a block, and the height of the lines in a block. The technique was tried on 52 Japanese documents and 21 English documents. No quantitative measure of performance was given.

Hinds et. al. (1990) sample a 300dpi document image by a factor of 4 and compute from it a burst image. A burst image can be obtained from the distance transform or the erosion transform of the document image using a 2 pixel vertical (horizontal) structuring element for portrait (landscape) mode images. The burst image selects only the column (row) relative maximum pixels of the erosion transform. They then compute the Hough transform of the burst image, incrementing each Hough bin by the value of the pixel in the burst image, providing its value is less than 25. The rotation skew of the image is then determined by searching the Hough parameter space for that bin having the largest accumulated value. Its angle is the rotation skew angle. They tested the technique on 13 document images. and correctly determined the rotation angle on all the images. The interline spacing on one document image was not correctly determined and the landscape/portrait mode was incorrectly determined on 5 document images.

Pavlidis and Zhou (1991) determine the geometric page layout by analyzing the white areas of a page by computing the vertical projection and looking for long white intervals from the projections. Then the column intervals are converted into column blocks, merging small blocks into larger blocks. Blocks are clustered according to their alignments and the rotation angle estimated for each cluster. The column blocks are then outlined. Finally, each block is labeled as text or non-text using features such as ratio of the mean length of black intervals to the mean length of white intervals, the number of black intervals over a certain length, and the total number of intervals. No performance results are given.

Baird (1992) discusses a computational geometry technique for geometric page layout by finding the maximal rectangles covering the white ares of the page. The rectanglar regions not covered by the maximal sized white rectangles then constitute regions which can then be classified as text or non-text.

Amamoto et. al. (1993) determine the geometric page layout by operating on the white space of the sampled document image. They open the white space of the sampled document image tith a long horizontal structuring element and open it with a long vertical structuring element. The union of these two openings then constitute the white space of the blocks. The blocks are then extracted from this white space. They decide that a block is a text block if the length of the longest black run length in the vertical and horizontal directions is smaller than a given threshold. A decision is made whether the writing is horizontal or vertical based on the number N_H of blocks whose width is greater than twice its height and the number N_V of blocks whose height is greater than twice its width. If $N_H > N_V$, then the decision is horizontal writing. Else vertical writing. Each block is then assigned a class label from the set: *text, figure, image, table*, and *separation line*. No performance results are given.

O'Gorman (1992) discusses what he calls the docstrum technique for determining geometric page layout. This technique involves computing the k-nearest neighbors for each of the black connected components of the page. Each pair of nearest neighbors has an associated distance and angle. By cluster the components using the distance and angle features, the geometric regions of a page layout can be determined.

Ishitani (1993) determines the rotation skew angle as that direction in which the variance of the complexity of the white-black transitions is greatest. Specifically, they define a set of lines for each angle. The difference between successive angles is .01 degree. Then they measure each line's complexity where complexity is defined as the number of white-to-black transitions along the line. The variance of the white-to-black transition counts is then determined. The angle which maximizes this variance is the estimated rotation skew angle. They report that this measure does not have difficulties with document pages which have large areas of non-text. The method was tested on 40 300dpi document images taken from magazines, newspapers, manuals and scientific journals. They report that the rotation angle was measured to within an accuracy of .12 degrees.

Hirayama (1993) develops a technique for determining the geometric layout structure of a document which begins by merging character strings into text groups. Border lines of blocks are determined by linking edges of text groups. Then blocks which have be oversegmented are merged and a projection profile method is applied to the resulting blocks to differentiate text areas from figure areas. Hirayama reports that on a data set of 61 pages of Japanese technical papers and magazines 93.3% of the text areas and 93.2% of the figure areas were correctly detected.

Ittner and Baird (1993) determine a geometric layout by doing skew and shear angle corrections, partitioning the page into blocks of text, inferring the text line orientation within each block, partitioning each block into text lines, isolating symbols within each text line, and finally merging the symbols into words. The rotation skew angle is determined by taking the projections of the centers of the connected components of the black pixels on the page at a given angle. The angle is iteratively updated to optimize the alignment without having to compute the projection over each possible angle. After rotating the image shear is corrected by a similar technique. They report an accuracy to within 3 minutes of arc indicating that the method fails perhaps one in one thousand images. Blocks are determined by the white space covering technique of Baird (1992). They report that on 100 English document image pages from 13 publishers and 22 styles, 94% of the layouts were correctly determined. The orientation of text lines in a block is determined from the minimum spanning tree of the connected components of the black pixels. The mode of the histogram of the directions of the edges in the minimum spanning tree is the orientation of the text lines in a block. Symbols in a text line are determined by taking the projection in an orthogonal direction to the text line. The projection profile is checked for a dominant frequency and the segmentation into characters is done from the projection profile with the knowledge of the dominant frequency. To determine the words in a text line, they determine a scalable word-space threshold for each text block separately. Then each text line is independently segmented to distinguish the inter-character spacing with the inter-word spacing.

Ankindele and Belaid (1993) determine a geometric page layout that permits blocks to be polygonal as well as rectangular. The determine the elongated white spaces in the document image and then determine the intersections of these white spaces. Points of intersection are candidate vertices for a polygonal block. The polygonal blocks are then extracted from the geometry of the intersection points. No performance results are given.

3 Logical Page Structure

Logical Page Structure involves determining the type of page (page classification) assigning functional labels to each block of the page, and ordering the text blocks according to their read order.

Esposito et. al. (1990) develop a symbolic learning approach and compare it with a statistical approach to accomplish page classification. They used a set of 231 pages and a seven page types and obtained 100% accuracy for 5 classes and 95% and 96% accuracy for the other 2 classes by a combined statistical and symbolic classification method.

Fujisawa and Nakano (1990) classified 81 Japanese Patent Disclosure Bulletin 200dpi document image pages into one of the three classes *front page, text page without figures*, or *text page with figures* with 100% accuracy. In a second test utilizing 106 Japanese Patent Application 200dpi document image pages, they had six classes: *separator sheet, front page type-1, front page type-2, claim page, text page*, or *figure page*. The recognition rate was 98%.

Tsujimoto and Asada (1990) assume that each block of the geometric page layout contains exactly one logical class. They organize the geometric page layout as a tree. Each new article in a document such as a newspaper begins with a headline which is in the head block. They find the paragraphs which belong to the head block by rules relating to the order of the geometric page layout tree and are able to assign logical structure labels of *title, abstract, sub-title, paragraph, header, footer, page number*, and *caption*. They worked on 106 document images and correctly determined the logical structure for 94 document images.

Visvanathan (1990) employs an $X-Y$ tree to represent the geometric layout and then employed a regular grammar scheme to label the document image blocks. Block labels included: *title, author, abstract, section titles, paragraphs, figure, table, footnote, footers* and *page numbers*. No performance results were given.

Fisher (1991) is an extension of Fisher (1990) and describes a rule based system to identify the geometrical and logical structure of document images. No performance results are given.

Ingold and Armangil (1991) describe a formal top-down method for determining the logical structure. Each document class has a formal description that includes composition rules and presentation rules. They have utilized the technique on legal documents. No performance results are given.

Chenevoy and Belaid (1991) use a blackboard system for a top-down method of logical structure analysis of a document image. The system is defined in a Lisp formalism and has a hypothesis management component using probabilities. No performance results are given.

Kreich et. al. (1991) describe a knowledge-based method for determining the logical structure of a document image. To obtain the blocks they search for the largest text blocks because these are the most characteristic elements in the document layout. The search consists of grouping together the connected components which are close enoough to each other. Once text blocks are determined, lines are found within each of the text blocks and words within the lines. The determination of document layout structure is based on interpreting documents and their parts as instance of hierarchically organized classes. They have defined over 300 classes for a document image and its parts. No performance results are given.

Derrien-Peden (1991) describes a frame-based system fo the determination of structure in a scientific and technical document image. The basis of this system is a macro-typographical analysis. The idea is that in a scientific and technical documents, changes of character size or thickness of type, white separating spaces, indentation etc. are used to make visual searching for information easier. So the technique searches for such typographical indications in the document and recovers its logical organization without any interpretation of its semantic content. The first step is the determination of the geometric page layout keeping a *part of* relationship between blocks. The logical structure determination removes running headers and footnotes and searches for the text reading order. Text blocks are then compared to logical models of classes and each text block is then assigned a class. No performance results are given.

Yamashita et. al. (1991) use a model-based method. Character strings, lines, and half-tone images are extracted from the document image. Vertical and horizontal field separators (long white areas or black lines) are detected based on the extracted elements, then appropriate labels are assigned to character strings by a relaxation method. Label classes included: *header, title, author, affiliation, abstract, body, page number, column, footnote, block* and *figure*. The technique was applied to 77 front pages of Japanese patent applications. They reported that the logical structure for 59 were determined perfectly.

Dengel (1993) discusses a technique for automatically determining the logical structure of business letters. He reports that on a test set of 100 letters, the recipient and the letter body could be correctly determined.

Saitoh et. al. (1993) determine logical layout with text block labels of *body, header, footer*, and *caption*. They tested the technique on 393 document images of mainly Japanese and some English documents. To characterize performance they measured the average number of times per document image an operator has to correct the results of the automatically produced layout. They report that on the average 2.17 times per image areas not suitable for output have to be discarded, .01 times per image mis-classified areas have to be correctly labeled, and 1.09 times per image does a text area have to be reset. With respect to text ordering they report that it required moving connections .47 times per image, on the average, making new connections .11 times per image, and re-assigning type of text .36 times per image.

4 Reflections

It is clear that many of the published papers give illustrative results and hardly any have their techniques tried out on significant sized data sets. It is certainly the case that since document understanding techniques will be moving to the consumer market, they will have to perform nearly perfectly. This means that they will have to be proved out on data sets whose document pages number in the thousands and that there must be suitable performance metrics for each kind of information a document understanding technique infers. Appropriate performance metrics for geometric page layout or logical page structure are not obvious. and they have hardly been discussed, Kanai et. al. (1993) being an exception. Yet each algorithm implicitly incorporates some kind of criterion to be optimized.

Publically available document image data sets whose correct geometric layout and logical layout are specified for each image in the data set are needed. Unfortunately, such data sets created by commercial suppliers of OCR, who are the natural suppliers of document understanding software, are kept proprietary. An example of one significant sized English document image data base publically available is that of Phillips et. al. (1993a,b). That data set has over 1000 English document images from scientific and technical articles which are labeled with their geometric layout and their logical structure.

It is interesting that although near perfect performance is required, very few of the document understanding related techniques develop a method which given the training data, theoretically optimizes the performance. This means that document degradation perturbation models and models to describe document format variations have not been formulated in a rich enough, simple enough, or complete enough way to permit their theoretical use.

Although some models for document degradation have been published, (Baird, 1993; Kanungo et. al. 1993) none has been validated. This is an important area to which more attention must be paid. For with random perturbation models that describe the range of variations and perturbations in document that might be encountered, it becomes possible to extend labeled document image data sets by successively degrading and perturbing and varying the document images in the data set and synthetically create larger data sets for the purpose of document image understanding algorithm development and testing. Even if such models are not completely validated they are useful since they can often be used to indicate where an existing document understanding algorithm is the weakest.

References

[1] O.T. Akindele and A. Belaid, "Page Segmentation by Segment Tracing," *2nd ICDAR*, Tsukuba, 1993, p341-344.

[2] N. Amamoto, S. Torigoe, Y. Hirogaki, "Block Segmentation and Text Area Extraction of Vertically/Horizontally Written Document," *2nd ICDAR*, Tsukuba, 1993, p739-742.

[3] H.S. Baird, "Background Structure in Document Images," *Advances In Structural and Syntactic Pattern Recognition*, World Scientific, Singapore, 1992, p253-269.

[4] H.S. Baird, "Document Image Defect Models and Their Uses," *2nd ICDAR*, Tsukuba, 1993, p62-67.

[5] D.J. Ittner and H.S. Baird, "Language-Free Layout Analysis," *2nd ICDAR*, Tsukuba, 1993, p336-340.

[6] D.S. Bloomberg, "Multiresolution Morphological Approach to Document Image Analysis," *1st ICDAR*, Saint-Malo, 1991, p963-971.

[7] R.G. Casey and G. Nagy, "Document Analysis – A Broader View," *1st ICDAR*, Saint-Malo, 1991, p839-849.

[8] Y. Chenevoy and A. Belaid, "Hypothesis Management for Structured Document Recognition," *1st ICDAR*, Saint-Malo, 1991, p121-129.

[9] A. Dengel, "Initial Leraning of Document Structure," *2nd ICDAR*, Tsukuba, 1993, p86-90.

[10] D. Derrien-Peden, "Frame-based System for Macro-typographical Structure Analysis in Scientific Papers," *1st ICDAR*, Saint-Malo, 1991, p311-319.

[11] F. Esposito, D. Malerba, G. Semeraro, "An Experimental Page Layout Recognition System For Office Document Automatic Classification: An

Integrated Approach For Inductive Generalization," *10th ICPR*, Atlantic City, 1990, p557-562.

[12] J.L. Fisher, "Logical Structure Descriptions of Segmented Document Images," *1st ICDAR*, Saint-Malo, 1991, p302-310.

[13] J.L. Fisher, S.C. Hinds, and D.P. D'Amato, "A Rule-Based System For Document Image-Segmentation," *10th ICPR*, Atlantic City, 1990, p567-572.

[14] H. Fujisawa and Y. Nakano, "A Top-Down Approach For The Analysis Of Documents," *Proceedings of the IAPR Workshop on Syntactic and Structural Pattern Recognition*, Murray Hill, New Jersey, 1990, p113-122.

[15] X. Hao, J.T.L. Wang, P.A. Ng, "Nested Segmentation: An Approach for Layout Analysis in Document Classification," *2nd ICDAR*, Tsukuba, 1993, p319-322.

[16] J. Higashino and H. Fujisawa, Y. Nakano, and M. Ejiri, "A Knowledge Based Segmentation Method for Document Understanding," *8th ICPR*, Paris, 1986, p745-748.

[17] S.C. Hinds, J.L. Fisher and D.P. D'Amato, "A Document Skew Detection Method Using Run-Length Encoding And The Hough Transform," *10th ICPR*, Atlantic City, 1990, p464-468.

[18] Y. Hirayama, "A Block Segmentation Method For Document Images with Complicated COlumn Structures," *2nd ICDAR*, Tsukuba, 1993, p91-94.

[19] R. Ingold and D. Armangil, "A Top-Down Document Analysis Method For Logical Structure Recognition," *1st ICDAR*, Saint-Malo, 1991, p41-49.

[20] Yasuto Ishitani, "Document Skew Detection Based on Local Region Complexity," *2nd ICDAR*, Tsukuba, 1993, p49-52.

[21] J. Kanai, T.A. Naratker, S.V. Rice, and G. Nagy, "Performance Metrics for Document Understanding Systems," *2nd ICDAR*, Tsukuba, 1993, p424-427.

[22] T. Kanungo, R.M. Haralick, I.T. Phillips, "Global and Local Document Degradation Models," *2nd ICDAR*, Tsukuba, 1993, p730-734.

[23] J. Kreich, A. Luhn, and G. Maderlechner, "An Experimental Environment for Model Based Document Analysis," *1st ICDAR*, Saint-Malo, 1991, p50-58.

[24] F. Lebourgeois, Z. Bublinski, and H. Emptoz, "A Fast and Efficient Method For Extracting Text Paragraphs and Graphics from Unconstrained Documents," *11th ICPR*, The Hague, 1992, p272-276.

[25] G. Nagy and S.C. Seth, "Hierarchical Representation of Opically Scanned Documents," *7th ICPR*, Montreal, 1984, p347-349.

[26] G. Nagy, S.C. Seth, S.D. Stoddard, "Document Analysis With An Expert System," *Pattern Recognition in Practice II*, E.S. Gelsema and L.N. Kanal editors, North Holland, Amsterdam, 1986, p149-159.

[27] L. O'Gorman, "The Document Spectrum for Bottom-Up Page Layout Analysis," *Advances In Structural and Syntactic Pattern Recognition*, World Scientific, Singapore, 1992, p270-279.

[28] T. Pavlidis and J. Zhou, "Page Segmentation by White Streams," *1st ICDAR*, Saint-Malo, 1991, p945-953.

[29] I.T. Phillips, S. Chen, and R.M. Haralick, "CD-ROM Document Database Standard," *2nd ICDAR*, Tsukuba, 1993a, p478-483.

[30] I.T. Phillips, J. Ha, R.M. Haralick, and D. Dori, "The Implementation Methodology for a CD-ROM English Document Database," *2nd ICDAR*, Tuskuba, 1993b, p484-487.

[31] T. Saitoh and T. Pavlidis, " Page Segmentation without Rectangle Assumption," *11th ICPR*, The Hague, 1992, p 277-280.

[32] T. Saitoh, M. Tachikawa, T. Yamaai, "Document Image Segmentation and Text Area Ordering," *2nd ICDAR*, Tsukuba, 1993, p323-329.

[33] Y.Y. Tang, C.Y.Suen, C.D. Yan, and M. Cheriet, "Docuyment Analysis and Understanding: A Brief Survey," *1st ICDAR*, Saint-Malo, 1991, p17-31.

[34] S. Tsujimoto and H. Asada, "Understanding Multi-articled Documents," *10th ICPR*, Atlantic City, 1990, p551-556.

[35] F.M. Wahl, K.Y. Wong, and R.G. Casey, "Block Segmentation and Text Extraction in Mixed Text/Image Documents," *Computer Graphics and Image Processing* Vol 20, 1982, p375-390.

[36] A. Yamashita, T. Amasno, H. Takahashi, and K. Toyokawa, "A Model Based Layout Understanding Method for the Document Recognition System," *1st ICDAR*, Saint-Malo, 1991, p130-138.

[37] C.L. Yu, Y.Y. Tand and C.Y. Suen, "Document Architecture Language (DAL) Approach to Document Processing," *2nd ICDAR*, Tsukuba, 1993, p103-106.