Research Journal of Recent Sciences _____ ISSN 2277-2502
Vol. **3(4),** 71-77, April **(2014)** *Res.J.Recent Sci.*

The Analysis of Connected Components and Clustering in Segmentation of Persian Texts

Askarpour S.[1], **Saberi Anari M.**[2], **Brumandnia A.**[3] **and Javidi M.M.**[4]
[1]Faculty Member of Technical and Vocation University, Kerman, IRAN
[2]Faculty Member of Technical and Vocation University,Yazd, IRAN
[3]Faculty member of Azad University, South Tehran Branch, Tehran, IRAN
[4]Faculty member of Shahid Bahonar University, Kerman, IRAN

Available online at: www.isca.in, www.isca.me
Received 8th December 2013, revised 13th February 2014, accepted 29th March 2014

Abstract

According to the application development computer in human life and increasing use of structured electronic documents and advantages of using them, the need to convert paper documents into their electronic format and use of image processing has been increased. Among researches that have been done in this field, we can point to the identification of the words in texts that comprehensive researches have been done in different languages such as : English, Japanese and Chinese. However, in Persian and Arabic languages, due to the complexity of these languages such as letters interconnection and various forms for letters according to their position in word, it is still need to research in this field. Segmentation is one of the most important steps in letter recognition system that it accuracy and speed is very important. Segmentation of Persian texts is the hardest since the specification of this language. In this study, we try to present a fast and efficient algorithm than same algorithms for segmentation of Persian documents with that help of connected components and clustering, we pay to identification and grouping of text and image areas. The users of this project are typical and we can use it as preprocessing steps of Optical Character Recognition systems. This research has been done on a collection of 100 scanned images of Persian newspapers and magazines with 300 dpi clarification and also it shows the simulation results with accuracy rate of %92.3 and significant speed than other approaches such as Voronoi Diagram.

Keywords: Page segmantation, connected components, clustering, persian text process, image processing.

Introduction

With development of computer application in all aspects of life providing ideas without paper and increasing growth of electronic documents utilization for saving, marketing and their better and faster processing have been caused many researches in identification and digitization of paper texts. Converting text image to text file is done by OCR system. Persian text segmentation is preprocessing step for converting text images to text files. First, in segmentation, text regions of an image are identified and in next step, the words of segmented regions convert to discrete characters and in last step, the characters are identified by an OCR Persian system. Accuracy in text segmentation causes that recognition steps are done better and faster.

Text segmentation is implemented by the up to down, down to up, synthetic and textural methods and it is still, one of the important researches in field of image processing and pattern recognition. The aim of this research is to provide a new and efficient method in Persian text segmentation. In the following, we review done tasks in segmentation and after presenting our algorithm, we compare our result with Voronoi Diagram.

A Review of done works: Some of the approaches of page segmentation or page analysis have been published in texts[1,2]. Some the approaches in image of a document consider the homogeneous regions (such as text, image and painting) as a region of the texture, then page segmentation is implemented with finding textural regions in gray images. Representative of this approach is the approach Jain and et al[3,4,5]. Some of the approaches of page segmentation act on background pixels processing by using white space to recognize homogeneous regions in an image[6-10].

These techniques are methods including :x-y tree[11,12,8], curve based on pixel layout[13], curve based on the layout of connected components[16] and trace of white space[10].

These approaches work by up to down[14,15,16] and down to up methods and also some of the techniques work by a combination of two above methods[17,18]. In Dakstran method[19], by using clustering K-Nearest Neighbor, groups the characters in to text lines and text blocks. In done work on identifying Persian language by Parhami and Tereghi[20], word segmentation to characters is done as below. First, input image is scrolling column by column. If a column of a segment only has black pixels and its size be close to font size. Then, we will have a potential segmentation column. So, a sequence of connection points obtained on Persian literature baseline. Their proposed technique uses a baseline that is an important property for

Persian and Arabic texts. In fact, where the thickness of baseline changes from its normal value, there will be the Junction point. Zang and et al[21] presented a technique that not only was based on structural characteristics between background and characters, but also, it was separated based on Arabic character properties. This technique whether a subword is just containing a character or not. Then it uses vertical histogram and carrying multiple rules for reaching to segmentation points. Yeen and et al[22] paid to separating the text line in Chinese handwritten documents. They used a text that had curvature and low distance between the lines and finally, line segmentation algorithm was presented based on clustering and same paragraph.

Hessan Al Ahmad and et al[23] showed a method of Arabic handwritten segmentation based on neural networks. Romeo parker and et al[24] published two methods that segmented continuous handwritten texts into characters. They segmented Arabic handwritten texts and continuous Latin texts of rows angle. They used the applied methods in Horizontal and vertical radiation and freeman chain code of laws. Rastegarpour and Shanbehzadeh (2007)[25] worked on segmentation of handwritten Persian digits in check. They used the improved algorithm (Vertical Radiation Curve) for segmentation of each row to words and sub words and used it for finding segmentation point and place of overlap and stick. Due to the limited numbers, they could perform the identification step with a collected database. Some worked by finding baseline method. Base line is a horizontal line that one line of a text locates on it, and its width is equal to font size and covers the maximum number of black pixels of that line text.

Pichoetez and et al[26] presented a method in that first, baseline was obtained as pieces of curves. Then, by using the best linear fit, baseline was estimated. In another method, by finding exertion center of connected components and using linear fit among these centers, this line is estimated. Hashemi and et al[27] first scrolled, the outer curve of the word for identifying Persian text and segmenting the nonbinding are raised characters and for most characters, they used moving window approach.

Text segmentation algorithm on Voronoi Diagram was proposed by Kise and et al[28] and it is considered an up-down algorithm. In first step, this algorithm extracts sample point for the borders of connected components by using sampling rate. Then, by using maximum threshold of noise region size, noise elimination is done.

After that, Voronoi diagram will be made by using obtained point form connected component's border. Voronoi edges passes from the middle of connected components. Finally, waste Voronoi edges are eliminated for obtaining the borders of document components. If, one edge has one of the their criterions, it considers as waste. The output of algorithm is including the regions that have been formed optionally by veronoi edges.

Azmi and Kabir[29] presented an algorithm for segmentation of Persian text with every fonts. They used conditional labeling rules and worked on upper contour instead of high vertical incision below the word. After pre processing step and identification of global baseline, we pay to contour survey. We recognize the segmentation point according to the grammar and correct them in post processing step. In order to find the font size, a text line is scanned column by column. The size of block pixels in these column that have the highest frequency, are considered as the font size. Motoa and et al[30] proposed an algorithm for segmentation of Arabic words to characters. They used mathematical morphological techniques. Safabakhsh and Adibi[31] applied a new technique for nastaliq style except finding local minimum of high contour technique. Boromandnia and Shanbehzadeh worked on segmentation of Persian text and a new wavelet transform has been applied in their proposed algorithm and extracted wavelet coefficients are used in identification of emphasizes horizontal edges. Irradiation of horizontal edges and their position on the baseline, provides segmentation points.

Proposed Method

The main goal of this article is to implement the segmentation phase in optical detection system that pays to identify and category text regions in images which has application in text steps of OCR. The base of this article is the utilization of analytical method and down to up segmentation for things classification in scanned images of Persian texts that as a result, text and image regions are identified. The inputs are scanned text images with proper quality. Processing is done on binary images and if the input image is colorful, it will convert to gray image and then binary image by finding a threshold. Design innovation is in identification and classification of text Persian texts indifferent points of image by clustering and connected components techniques.

The proposed method is resistant against noise and curvature than obvious methods and also it has improved the speed and accuracy of present algorithms.

Description of proposed steps

Preprocessing step: In most identification character methods, preprocessing is done. In this step, for working on image of a text in computer, we should enter it on computer and clear. Preprocessing operation includes image scan, binary optimization, noise elimination and curvature correction. In fact, in this steps input image clears by image processing techniques and gets ready for segmentation steps.

Image Readying: In this step, paper documents are scanned by scanner with 300dpi clarity and entered on computer. Produced numbers information on computer are saved in matrix by scanners in that each component shows the brightness of the corresponding point in image. Matrix size depends on document

image size, clarity degree of scanner (dpi) and dept (the numbers of gray or colored levels).

Binary optimization of image: In order to decrease the memory consumption and corresponding processing time, it's better to encode the pixels as 0,1 and black and white images are processed instead of gray and colored images. Mean while, the process of converting gray scale image or colored image to white or black image is called binary optimization or thresholding because the obtained image is that image which its corresponding matrix is only binary number 0 and 1. A binary image is as black and white, and include all the necessary information according to numbers, position and object's shape.

So, it is containing lower volume information than gray and colored image because in obtained binary image, unnecessary information related to background is eliminated and just data, necessary information related to background and main image are method for binary optimization of image is such that first, a threshold value is chosen by a method and then for all the image pixels with brightness lower than threshold value, amount and for all pixels with brightness more than threshold value, a mount is considered. Global threshold value, with regardless of input image class, is a value in [0, 1].

Correction of image curvature: Estimation of inclination angle and image correction of a document page is an important principle for analyzing the text and optical character identification many methods have been presented in digital texts and images for curvature correction and for more study, you can refer to references. Some of these methods are working on pixels' level, finding baseline, connected components' level, nearest neighbor and so on.

Noise elimination: All the image acquisition processes result in different kinds of noise mean while, there is no ideal way for the absence of noise. Noise is predictable and is exactly measurable by a noisy image. Two kinds of noise have been defined: signal-independent noise and signal-dependent noise. Signal-independent noise adds a random set of gray level to image pixels and it is statistically independent of image data. In signal-dependent noise, each value of a point in image is a function of gray levels.

When the text is on a checkered page, with scanning the main page, a page is produced that structural or alternative noise is distributed in it. In these cases, Fourier transform has application and used for determining the peak frequency. Noise is related to one of the peak and we can virtually eliminate noise by elimination of these values from frequency domain and convert the image to a systematic image.

Finding the connected components step and drwing a surrounded rectangles: finding and marking tag of image connected components. In this article, we are seeking for different regions in scanned images. These images are containing many objects such as text, image, table and soon. We should find the connected components of the image. The algorithm of connected component's marking tag categorize all the connected components of an image based on clustering technique and criterion of proximity and neighboring. If two groups were neighbor with each other(had a neighbor pixel), they will be merged with each other and in first, all the clusters are single pixel and all the present points in a same , receive a same tag.

The analysis of connected components: Extraction of connected components in a binary image is an important and necessary task for many of image analysis applications. The most time-consuming in our segmentation algorithm and other same algorithm is related to clustering and marking tag of connected components because in this step, individual image pixels should be surveyed and processed[32]. After identification of foreground from background, all the pixels that are connected with each other and aren't connected with other pixels are identified categorized and marked as the connected components (figure 1). Pixel's numbers and tags of a component are saved in array matrix of that image and in corresponding elements of each pixel: Then, the algorithm of each component is showed by a pixel that is in its exertion center.

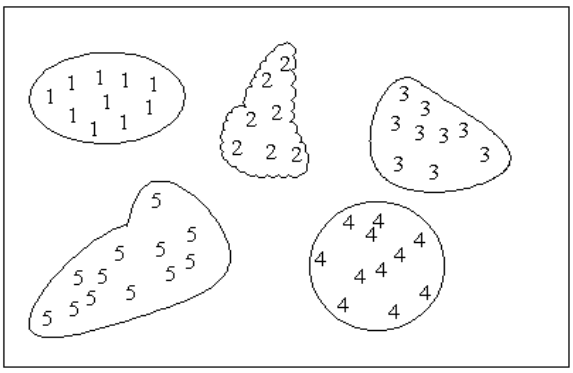

Figure-1
Marking tag of connected component's pixels

According to the approach of connected components analysis, we can categorize the pixels of image to connected components and then convert and merger the connected components to surrounded rectangles and surrounded rectangles to graphic text lines and graphic text lines to region blocks.

In next section, we explain more details of this analysis by the help of surrounded rectangle. Drawing surrounded rectangle for each connected component by finding three important parameters of a connected component, we can obtain its surrounded rectangle. In connected component of figure 2, we obtain the coordinate of the points (x, y), H, W for each component. H is the transverse distance of its last point and W is longitudinal distance of its last point. We obtain the surrounded rectangle of this section by drawing lines between the points A,B,C,D.

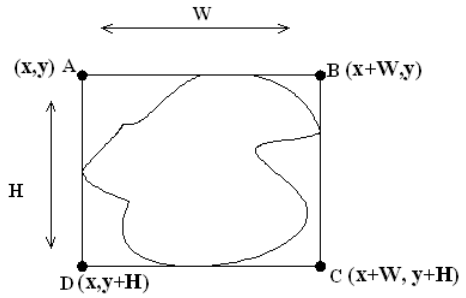

Figure-2
Drawn surrounded rectangle for a connected component

C={ c_i } is a set of connected components. The surrounded rectangle of each connected component, c_i, is identified by the coordination of highest point in left and lowest point in right. Two points with coordination $((X_u(C_i),Y_u(C_i)),(X_l(C_i),Y_l(C_i))$ are somehow that there are $X_u(C_i) < X_l(C_i)$ and $Y_u(C_i) < Y_l(C_i)$ [15].

Merge step: Some of the surrounded boxes and connected component are inside each other. In these cases, if these components were in text, they would related to a word or binary words and if they were in tables or images, they will be a retail of larger table or image.

Therefore, we merge and categorize them with each other to make the output better for next identification step for this task, we use the ideas and techniques of clustering. First we consider each component with an independent cluster and with a new criterion, we merge the distance between surrounded boxes of each component and make a new component with a new surrounded box.

The distance of surrounded boxes and horizontal overlap: The definition of distance between two surrounded boxes for classification and clustering is defined by Simon[33]. The distance between two connected component C_i, C_j or two objects O_i, O_j is stated as the distance between their surrounded boxes. We define a typical object O_i, that can be the connected component of the C_i. In left and up points and right down points, surrounded box of this component is as $((X_u(O_i),Y_u(O_i)),(X_l(O_i),Y_l(O_i))$. Coordination of X, Y points of this surrounded box is as $X_u(O_i) < X_l(O_i)$ and $Y_u(O_i) < Y_l(O_i)$. Width is defined as $W = X_l - X_u$ and height is defiend as $H = Y_l - Y_u$ that they are as the vertical size and horizontal sizes of an object, respectively.

Now, according to figure 3, we define the vertical and horizontal distance between two objects base on rules 1 and 2:

$$D_x(O_i,O_j) = \max[X_u(O_i),X_u(O_j)] - \min[X_l(O_i),X_l(O_j)] \quad (1)$$

$$D_y(O_i,O_j) = \max[Y_u(O_i),Y_u(O_j)] - \min[Y_l(O_i),Y_l(O_j)] \quad (2)$$

We can understand that if $D_x(O_i,O_j) < 0$, then object O_i and O_j have overlap horizontally and in the same way, we can define vertical overlap.

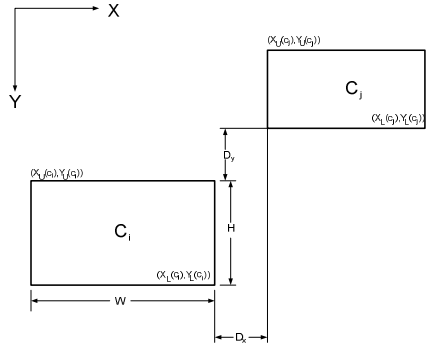

Figure-3
Horizontal and vertical distance of a surrounding box

Horizontal merge of components: After finding horizontal distance between surrounding boxes, two connected components are categorized with other if they have overlap or $D_x(C_i,C_j)$ is lower than T_d. T_d is a parameter that typical distance of words in a text and image is estimated based on it. In fact, for horizontal merging in a connected components set of C={C_i}, we state that both connected components C_i and C_j have overlap in a near horizontal distance of T_d if :

$$D_x(C_i,C_j) < T_d \quad (3)$$

If both connected components C_i and C_j are applies in equation 3, then, C_i and C_j are merged with each other and in a new cluster, the highest left points (u) and the lowest right points (l) are calculated as bellow:

$$X_u = \min[X_u(c_i),X_u(c_j)] \quad ,X_l = \max[X_l(c_i),X_l(c_j)] \quad (4)$$

$$Y_u = \min[Y_u(c_i),Y_u(c_j)] \quad ,Y_l = \max[Y_l(c_i),Y_l(c_j)] \quad (5)$$

For merging, surrounded boxes of connected components are saved in a linked list. So, each element of linked list includes the coordination of a surrounded box. Then, surrounded boxes that should be merged, are eliminates form linked list and after new coordination calculation for new surrounded box by rules 4 and 5, the surrounded box of new component is written in linked list.

Vertical merger of surrounded boxes: After finding vertical distance between surrounded boxes of two connected components, they are categorized with each other if they have overlap or $D_y(b_i,b_j)$ is lower than T_b. T_b is a parameter that distance of text lines in image is estimated based on it. In fact, for vertical merge, some of surrounded boxes of B{b_i} are categorized. For a set of given surrounded boxes of B={b_i}, we state that b_i, b_j have overlap in a near vertical distance if:

$$D_y(b_i, b_j) < T_b \tag{6}$$

If both surrounded boxes b_i, b_j are applied in equation 6, then b_i, b_j are merged with each other and in new cluster, new coordinations of highest left points(u) and lowest right points(l) are calculated as below:

$$X_u = \min[X_u(b_i), X_u(b_j)] \quad , X_l = \max[X_l(b_i), X_l(b_j)] \tag{7}$$

$$Y_u = \min[Y_u(b_i), Y_u(b_j)] \quad , Y_l = \max[Y_l(b_i), Y_l(b_j)] \tag{8}$$

For merging, surrounded boxes of B={b_j} are saved in a linked list. So, each element of linked list includes the coordination of a surrounded box. If T shows the linked list, surrounded boxes of B={b_j} are merged with each other. The result of algorithm's output of merging connected components has been showed in figures 4 and 5.

Figure-4
Surrounded rectangles of connected component in an input image

Figure-5
The output of merging surrounded rectangles and connected components

In figures 7,6, the applied results on a newspaper image have been shown.

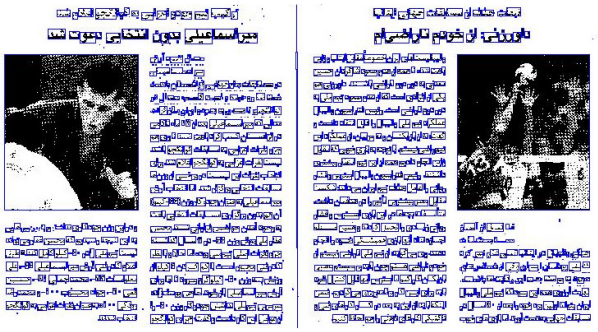

Figure-6
The output of connected component's algorithm and clustering

Figure-7
The output of proposed algorithm

The comparison of applied clustering techniques with some clustering techniques

First, applied techniques consider all the pixels and components as a single- element clusters and then emerge them so far as is necessary and it is like compressed technique. And also since, it regards to the distance between components, it is similar to single link clustering method. Innovative clustering method for this application, merges two components (with comparison of their surrounded boxes' distance) if they were close and gains the coordination of new surrounded box based on presented rules. So ,we can see that the calculation for obtaining the distance between components and their merged surrounded boxes' coordination is not much. As a result, for text segmentation, this clustering method is proper due to its low calculations. In single link clustering method, with many initial clusters, after merging two clusters, it needs more calculations for obtaining new distance of clusters with new merged clusters because the number of groups is high. However, in Text segmentation we deal with many pixels and initial same paragraph components.

In Single link method, we should look for the nearest distance for identifying the distance between clusters in order to obtain the distance between two clusters. This operation needs more time due to many pixels of a component. Calculations became very much for finding distances[34,35].

However, in applied method of this algorithm, the elements of each cluster are near each there and in a surrounded rectangle and for calculating the distance between two cluster, it is only enough to obtain the distance between their surrounded boxes and it won't need to calculate the distance between internal elements of clusters. Here, the similarity between clusters is equal to the distance between two surrounded box and if this is lower than zero or is a certain amount, we put them in a new cluster. Also, in this method, it dose not need to calculate and update the distance of each cluster with all clusters for clustering but also, the calculation of each cluster distance from clusters around it is enough. In complete link clustering method, the criterion of cluster's similarity is the distance between the farthest elements of two clusters. That the criterion of farthest neighbor is not proper for finding connected components and their merging in our method[36-39].

Experimental Results and Evaluation of Performance and its efficient

Proposed segmentation algorithm was tested after implementation on 100 images of Persian and local newspapers and magazines with 300dpi clarity. The tested images were containing text and image quantitative evaluation of this algorithm's performance on 100 images has been given in table 1.

So, accuracy rate of proposed algorithm is suitable for Persian text segmentation and comparison with other algorithms, it has better results. This algorithm has been implemented and tested in Matlab 7.10, Xps Dell system with characteristics cpu: 2.4 GHZ, RAM:512 MB consumed time for recognizing and marking tag of connected components of each image is average 1.7 seconds and for finding surrounded boxes and their merger, it was 0.32 seconds. As a result, total average consumed time was 2.02 second for proposed segmentation method in tested image that it was lower than same algorithms. We have compared the page segmentation algorithm with Voronoi Diagram algorithm and proposed segmentation algorithm of connected components for 100 scanned images of Persian magazines and the result of comparison have been given in tables 2 and 3.

Conclusion

Implementation of post processing step: in the end of work, we can survey proper methods such as strong database and accuracy of identification operation in words and eliminate the probable problems. Using the segmentation method with identification : if it performs with identification segmentation and segments again, if it is not true, we can increase accuracy to 100. Standardization and expansion of databases: word databases have not been completed in Persian language yet, and with improving and complementation of these databases, we can use it in all identification steps.

Using and expansion of this method for English handwritten texts : this method can be developed for English and Persian handwritten texts.

The speed of finding step and marking tag of same paragraph comment, in page text segmentation phase, needs the most time. However, new algorithms including what we used, decreased this time, if we can decrease the time of this step, segmentation time will be reduced to half and will have a suitable effect on OCR systems.

Table-3
The comparison of average time of implementation measured in second in voronoi segmentation algorithm and proposed method

Image condition	With curvature	Without curvature
Segmentation method		
Voronoi segmentation	3.6	3.5
Proposed algorithm	2.12	2.02

References

1. O'Gorman, L. and R. Kasturi, Document Image Analysis, Los Alamitos, California: IEEE computer Society Press, **(1995)**

2. Haralick, R. Document Image Understanding: Geometric and Logical Layout. in Proc. IEEE Conf. Computer Vision and Pattern Recognition **(1994)**

3. Jain, A. and Y. Zhong, Page Segmentation Using Texture Analysis. Pattern Recognition, 29, 743-77 **(1996)**

4. Jain, A. and K. Karu, Learning Texture Discrimination Masks. IEEE Trans, *Pattern Analysis and Machine Intelligence*, **18**, 195-20 **(1995)**

5. Jain A. and Bhattacharjee S., Text Segmentation Using Gabor Filters for Automatic Document Processing, Machine Vision and Applications, **5**, 169-184 **(1992)**

6. Ittner D. and Baird H., Language-Free Layout Analysis. in Proc. Second Int'l Conf. Document Analysis and Recognition, Tsukuba, Japan **(1993)**

7. Baird H., Anatomy of a Versatile Page Reader. in Proc. IEEE. **(1992**)

8. Antonacopoulos A. and Ritchings R., Flexible Page Segmentation Using the Background. in Proc. 12th Int'l Conf. Pattern Recognition, Jerusalem **(1994)**

9. Amamoto, N., S. Torigoe, and Y. Hirogaki. Block Segmentation and Text Area Extraction of Vertically/Horizontally Written Document. in Proc. Second Int'l Conf. Document Analysis and Recognition. Tsukuba, Japan **(1993)**

10. Akindele, O. and A. Belaid. Page Segmentation by Segment Tracing. in Proc. Second Int'l Conf. Document Analysis and Recognition. Tsukuba, Japan **(1993)**

11. Nagy, G. and S. Seth. Hierarchical Representation of Optically Scanned Documents. in Proc. Seventh Int'l Conf. Pattern Recognition, Montreal **(1984)**

12. Krishnamoorthy M., et al., Syntactic Segmentation and Labelingof Digitized Pages From Technical Journals. IEEE Trans. Pattern Analysis and Machine Intelligence, **15**, 743-747 **(1993)**

13. Pavlidis, T. and J. Zhou. Page Segmentation by White Streams. in Proc. First Int'l Conf. Document Analysis and Recognition. Saint-Malo, France **(1991)**

14. Ingold, R. and D. Armangil. A Top-Down Document Analysis Method for Logical Structure Recognition. in Proc. First Int'l Conf. Document Analysis and Recognition. Saint-Malo, France **(1991)**

15. Fujisawa H. and Nakano Y., A Top-Down Approach for the Analysis of Documents. in Proc. 10th Int'l Conf. Pattern Recognition. Atlantic City, N.J **(1990)**

16. Chenevoy Y. and A. Belaid. Hypothesis Management for Structured Document Recognition. in Proc. First Int'l Conf. Document Analysis and Recognition, Saint-Malo, France **(1991)**

17. Liu J., et al. Adaptive Document Segmentation and Geometric Relation Labeling: Algorithms and Experimental Results. in Proc. 13th Int'l Conf. Pattern Recognition : Vienn **(1996)**

18. Esposito F., Malerba D. and Semeraro G., A Knowledge-Based Approach to the Layout Analysis. in Proc. Third Int'l Conf. Document Analysis and Recognition, Montreal **(1995)**

19. O'Gorman, L., The Document Spectrum for Page Layout Analysis. IEEE Trans. Pattern Analysis and Machine Intelligence. **15**, 1162-1173 **(1993)**

20. Parhami B. and Tereghi M., Automatic recognition of printed Farsi Text. Pattern Recognition **(1981)**

21. Zheng L., H. A.H., and X. tang, A new algorithm for machine printed Arabic Character segmentation, pattern Recogniting Letters. 2(15), **(2004)**

22. Yin, F. and Cheng-LinLiu, Handwritten Chinese text line segmentation by clustering with distance metric learning, pattern recognition, **42**, 3146-3157 **(2009)**

23. Ahmad H.A. and Zitar R.A., Development of an efficient neural-based segmentation technique forArabic handwriting recognition, *Pattern Recognition*, **43**, 2773-2798 **(2010)**

24. Romeo-Parker K., Miled H. and Lecourtier Y., A New Approach for Latin/Arabic character segmentation, in in Proc. International Conference on Document Analysis and Recognition **(1995)**

25. Rastegarpour M. and J. Shanbezadeh, OFF-Line Hand-written Farsi/Arabic Word segmentation into subword under Overlapped or Connected Conditions. in The Springer Proceeding of International Workshop on Advances in pattern Recognition IWAPR2007. University of Loughborough, Plymouth-UK **(2007)**

26. Pechwitz M. and V. Marger, Baseline estimation for Arabic handwriting Recognition **(2002)**

27. Hashemi M.R., O. Fatemi, and R. safavi. Persian Cursive script Recognition. in proc. International conference on Document Analysis and Recognition **(1995)**

28. Kise K., Sato A. and Iwata M., Segmentation of Page Images Using the Area Voronoi Diagram. Computer Vision and Image Understanding, **70(3),** 370-382 **(1998)**

29. Azmi R. and Kabir E., A new segmentation technique for omnifont Farsi Text. Pattern Recognition Letters, 22 **(2001)**

30. Motawa D., A. Amin, and R. Sabourin ,segmentation of Arabic Cursive Script, in IEEE. **(1997)**

31. Safabakhsh R. and P. Adibi, Nastaaligh Hand Written Word Recogntion Using A Continuous-Density Variable-Duration Hmm. The Arabian Journal for Science and Engineering, 30(Number 1 B) **(2005)**

32. Jain R., Kasturi R. and Schunck B.G., Machine vision: McGraw-Hill, Inc **(1995)**

33. Simon A., Pret J. and Johnson A., A Fast Algorithm for Bottom-Up Document Layout Analysis, IEEE Trans. Pattern Analysis and Machine Intelligence, **19,** 273-276 **(1997)**

34. Kovács F., Legány C. and Babos A., Cluster Validity Measurement Techniques, Department of Automation and Applied Informatics, Budapest University of Technology and Economics **(2003)**

35. Keller, F., Clustering, in Tutorial Slides, Computer University Saarlandes, **(2003)**

36. Muhammad Altaf Khan, Islam S., Murad Ullah, Sher Afzal Khan, G. Zaman, Muhammad Arif1 and Syed Farasat Sadiq, Application of Homotopy Perturbation Method to Vector Host Epidemic Model with Non-Linear Incidences*, Research Journal of Recent Sciences,* 2(6), 90-95 **(2013)**

37. Farooq Ahmad, Sher Afzal Khan, Ilyas Fakhir and Yaser Daanial Khan, A Survey on Linear Algebraic Approaches for the Analysis of Petri Net based Models, *Res. J. Recent Sci.*, **2(5),** 21-28 **(2013)**

38. Panah Amir, Enhanced SLAM for a Mobile Robot using Unscented Kalman Filter and Radial Basis Function Neural Network, *Res. J. Recent Sci.,* **2(2),** 69-75 **(2013)**

39. Belsare Satish and Patil Sunil, Study and Evaluation of uses behavior in e-commerce Using Data Mining, *Res. J. Recent Sci.,* 1(**ISC-2011),** 375-387 **(2012)**