Journal of Theoretical and Applied Information Technology
20th December 2013. Vol. 58 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

SOUTH INDIAN TAMIL LANGUAGE HANDWRITTEN DOCUMENT TEXT LINE SEGMENTATION TECHNIQUE WITH AID OF SLIDING WINDOW AND SKEWING OPERATIONS

SUNANDA DIXIT[1], Dr.H.N.SURESH[2]
[1]Assistant Professor
Department of Information Science and Engineering
Dayananda Sagar college of Engineering, Bangalore 560078.
[2]Professor Dept. of Instrumentation Technology
Bangalore Institute of Technology, Bangalore-560004, Karnataka.
[1]E-mail: sunandadixit0776@gmail.com , sunanda.mah@gmail.com

ABSTRACT

In document image analysis, Text line segmentation is one of the key components. The segmentation logic presents essential information about skew correction, zone segmentation, and character recognition. The method of document image segmentation into text lines for printed text has seen numerous contributions from fellow research scholars, yet there is scope for tremendous improvement. The key challenges for handwritten document are due to writer movement, the inter-line distance changeability and incoherent distance between the components that may differ. These may be directly by segments, or curved. The area of handwritten segmentation has seen few models; very few of the research paper are proposed for Text line skew segmentation model and hence the stimulus of handwritten south Indian languages. Consequently, a better text line segmentation technique for south Indian Tamil language is proposed in this paper. The processing of Tamil language is very crucial factor because the Tamil letters are in crucial shapes and it is harder to segment the touching lines and letters from the Tamil image documents. The challenges present in Tamil language process and the existing text line segmentation methods has been improved by our proposed method, which utilizing two major techniques namely, sliding window and adaptive histogram equalization. Our proposed text line segmentation technique initially performs the preprocessing process and these preprocessed document images are given to the adaptive histogram equalization. During the histogram equalization process, the document images text characters are enhanced to view the characters more accurately. The enhanced image text lines are segmented by utilizing the sliding window operation. For accurate line segmentation, the skewing operation is performed on the line segmented result images. The implementation result shows the effectiveness of proposed technique, in segmenting the handwritten text lines from the input document. The performance of the proposed technique is evaluated by comparing the result of proposed technique with the conventional text line segmentation technique. The result shows that our proposed technique acquires high-quality text line segmentation DR, RA and F-Measure values for the number of testing documents in comparison with the conventional technique.

Keywords: *Line Segmentation, Sliding Window, Optical Character Segmentation, Adaptive Histogram, Skewing*

1. INTRODUCTION

Text line segmentation is one of the main elements in document image analysis. It offers decisive information for skew correction, zone segmentation, and character recognition [1]. Freestyle handwritten text lines gives an important challenge [2], even though text line segmentation technique for machine printed or hand-printed documents have been generally seen as a worked out problem. In document image analysis, handwritten text line segmentation is still regarded to be a most important challenge. In a handwritten text document analysis processing pipeline, it would pursue image binarization and page segmentationas a first step before word and character segmentation, character recognition etc. [3] Since it is in the commencement of a pipeline of processing, it is terribly essential to minimize faults so that next steps for pipeline get precise input.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Line segmentation is a method in which the successive lines are extorted or divided from each other to form a text [4]. Line segmentation can be broadly classified as two kinds: i) Typed text; ii) Handwritten text [5]. In typed text, the spcing between the lines is significant while the elements of the two successive lines may be contacted or overlapped in the hand written text, creating the problem more difficult to interline distance changeability and the base line skew changeability. Thus the line segmentation of unconstrained hand written text is not easy [6].

India is a multi-lingual and multi-script country, containing eighteen authorized languages [7]. Indian languages have a more sophisticated notion of a character unit or akshara that forms the Fundamental linguistic unit. This common alphabet contains 33 consonants and 15 vowels in common practice. Additional 3-4 consonants and 2-3 vowels are used in specific languages or in the classical forms of others. This difference is not very significant in practice. In difference of the skew angle among text lines or along the same text line, survival of overlapping or touching lines, changeable character size and non-Manhattan layout are the confronts of text line extraction [8]. Owing to high changeability of writing styles, scripts, etc., techniques that do not employ any former knowledge and adapt to the properties of the document image, as the offered, would be more tough. Line extraction methods may be classified as projection based, grouping, smearing and Hough-based [9]. Few of the well known techniques for text line segmentation are almost classified as follows. Smearing methods: short white runs are packed with black pixels proposing to create large bodies of black pixels, which will be regarded as text line areas [10]. Hough based classification have been used framework for tracking the type of object is not restricted to specific classes or categories of multi-lingual and multi-script.

The process of removing objects of interest from an image to text is known as Segmentation. Distinguishing lines is the initial step in segmentation [11]. The following steps are identifying the words in every line and the individual characters in every word. The first step in segmentation is detecting lines. The next subsequent steps are detecting the words in each line and performance of any recognition system. This is an essential step of OCR systems as it removes meaningful areas for investigation [12]. This step effort to decay the image into classifiable divisions called character. With touching and overlapping elements smearing techniques cannot contract fit. Horizontal projections: a vector having the sums of every image line is produced. The local minima of that vector are supposed to be the protrusion of white areas in between lines, and the image is fragmented consequently [13] [14].

The difference between related shaped components is the most challenging part of the Indian handwritten character identification [15]. Kannada is one of the main Dravidian languages of Southern India, one of the initial languages proved epigraphically in India. The writing has 49 characters in its alphasyllabary and is phonetic. Tamil is one of the oldest languages in the world and is a Dravidian language [16]. The Tamil writing has 10 numerals, 12 vowels, 18 consonants and five grantha letters. Telugu is a Dravidian language and has the third main well-liked writing in India [17]. There are 10 numerals, 18 vowels, 36 consonants, and three twofold symbols. Malayalam is a Dravidian language, and has the eighth most well-liked writing in India. Still, Malayalam has its own writing [18]. The writing has 16 vowels, 37 consonants and 10 numerals. An incredibly small difference among two characters or numerals leads to identification difficulty and a convinced degree of identification precision [19]. The method of writing the characters is extremely dissimilar, as they come in different sizes and shapes. The similar numeral may get dissimilar shapes, and on the contrary, two or more different numerals of writing may receive a similar shape [20].

The paper content is organized as follows: The research works associated to the handwritten manuscripts text line segmentation and Problem Statement is specified in Section 2. The handwritten documents text line segmentation by our proposed method is described in section 3. The experimental result and ending of this paper is specified in Section 4 and 5.

2. RELATED WORKS

For handwritten Gujarati numbers, Apurva A. Desai *et al.* [21] have explained with an optical character recognition (OCR) system. One may come across so much of work for Indian languages like Hindi, Kannada, Tamil, Bangala, Malayalam, Gurumukhi etc, but Gujarati was a language for which barely any work was observable particularly for handwritten characters. At this point in this suggested work a neural network was advised for Gujarati handwritten digits recognition. For categorization of digits, a multi layered feed forward neural network was recommended. Ahead

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

of their categorization, thinning and skew-correction were moreover made for preprocessing of handwritten numerals. For Gujarati handwritten digit recognition this work had attained roughly 82% of success rate.

A multi-stream strategy for off-line handwritten word identification has been offered by Yousri Kessentini *et al.* [22]. For the incorporation of various sources of data, the multi-stream paradigm presents an interesting structure and was compared to the normal grouping strategies namely fusion of representations and fusion of decisions. They examined the expansion of 2-stream approach to N streams (N=2... 4) and explore the development in the identification performance. The computational charge of this expansion was conversed. The multi-stream structure develops the identification performance in both cases. With 2-stream approach, the top identification performance was 79.8%, in the case of the Arabic script, on a 2100-word lexicon containing 946 Tunisian town/village names. The suggested approach reaches an identification rate of 89.8% by means of a lexicon of 196 words in the case of the Latin script.

Multilingual character identification system and their test out performances on standard database has been developed by V. N. Manjunath Aradhya *et al.*[23], there was yet room to develop the identification charge by progressing better features. In this projected method, they offered a multilingual character recognition system for published South Indian scripts (Kannada, Telugu, Tamil and Malayalam) and English texts. South Indian languages are most well known languages in India and all around the world. Based on Fourier transform and principal component analysis (PCA), the suggested multilingual character recognition was which are two generally applied methods of image dealing out and identification. In the region of pattern identification and computer vision, PCA and Fourier transforms are typical feature extortion and data representation methods extensively used. Experimental results demonstrate the presentation over the data sets regarded at this point.

Mamatha Hosalli Ramappa *et al.* [24] have suggested a method Optical character recognition (OCR) submits to a course of generating a character input by optical means, like scanning, for identification in following steps by which a printed or handwritten passage. A generic character recognition system has dissimilar steps like noise removal, skew detection and correction, segmentation, feature extraction and classification. In the OCR process, results of the presently stages could influence the presentation of the following stages. The skew finding and correction and segmentation act an essential part to compile the results of the following stages more precise. In this suggested technique, they have planned ideas for skew recognition and correction, segmentation of handwritten Kannada text by means of bounding box technique, Hough transform and contour detection correspondingly. A standard segmentation rate of 91% and 70% for lines and words was attained respectively.

OCR system, which depends on the segmentation algorithm being used, has been offered by Mamatha H R*et al.* [25]. Fragmentation of handwritten manuscript of some Indian languages like Kannada, Telugu, It consists of vowels, consonants and compound characters. A few of the characters may go beyond as one. Improvement of OCR tools in Indian languages was still a continuing practice in spite of numerous booming works in OCR all over the world. Character segmentation acts an essential part in character recognition since imperfectly segmented characters are improbable to be accepted properly. In this suggested technique, a segmentation plan for fragmenting handwritten manuscript by morphological operations and projection profiles was recommended. Practice of the morphology prepared removing text lines competent by a standard extraction rate of 94.5%. Because of the changeable inter and intra word gaps an average fragmentation rate of 82.35% and 73.08% for words and characters correspondingly was attained.

The Problem Statement

Section 2 reviews about the recent research works related to the segmentation of handwritten text of south Indian languages. The review has explored the research inclinations in the experimentation of the hand-written documents. Most of the review works have been performed using histogram techniques, Optical Character Recognition (OCR) system, which can improve the thinning and skew-correction of handwritten text line before their segmentation. We are inspired by the idea that a text image can be strip into images along the white gaps in between text lines. For all these tasks, a major step is document image segmentation into text lines. Because of the low quality and the complexity of these handwritten south Indian languages (background noise, artifacts due to aging, interfering lines), automatic text line segmentation remains an open research field. The inter-line distance variability and inconsistent distance between the components may vary due to

Journal of Theoretical and Applied Information Technology
20th December 2013. Vol. 58 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

writer movement. It may be straight by segments, or curved. A less number of researchers have contributed to develop such type of model for the hand-written documents. Our research paper considers the base on Text line skew segmentation model and proposes a novel technique of handwritten south Indian languages text line segmentation.

3. PROPOSED METHOD FOR SOUTH INDIAN TAMIL LANGUAGE HANDWRITTEN TEXT LINE SEGMENTATION TECHNIQUE

Our proposed handwritten text line segmentation technique for Tamil language segments the text lines from the input handwritten document images by using the sliding window and skewing operations. The proposed technique mainly comprised of four stages namely, (i) Preprocessing (ii) Adaptive Histogram Equalization (iii) Sliding window based line segmentation and (iv) Skewing Operation. These four stages are consecutively performed and the handwritten documents text lines are segmented more accurately and are discussed in Section 3.1, 3.2, 3.3 and 3.4 respectively. Structure of our proposed handwritten Tamil language text line segmentation technique is illustrated in Figure. 1.

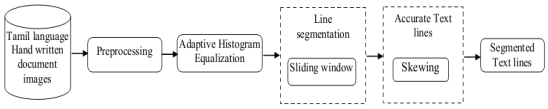


Figure 1: Structure Of Our Handwritten Tamil Language Text Line Segmentation Technique Based On Sliding Window And Skewing Operations

3.1 Preprocessing

Preprocessing process ensures to attain high segmentation accuracy. The input handwritten document image is named as $d(m \times n)$ and given to the preprocessing process. The process of preprocessing is shown in Figure. 2. Initially, the given document image $(d_i)_{mn}; i = 1,2,\cdots D, m = 1,2,\cdots M, n = 1,2,\cdots N$ where i represents the number of document images and M, N represents the row and column of the image. The grayscale converted image $(d_i)_{mn}$ is named as $G(d_i)_{mn}$. Next, this $G(d_i)_{mn}$ image is given to the binarization process. Binarization is a technique, where the gray scale images are converted into binary images. Binarization separates the foreground (text) and background information. The process of binarization is stated as follows,

$$\overline{G(d_i)_{mn}} = \begin{cases} 1; G(d_i)_{mn} > T \\ 0; otherwise \end{cases} \quad (1)$$

In Egu. (1), $\overline{G(d_i)_{mn}}$ is the binary image and T is a threshold value. If the image $G(d_i)_{mn}$ intensity value is greater than the given threshold value means we change the pixels values into 1, otherwise we change into 0. Afterward, the non textual area is removed from the binary image $\overline{G(d_i)_{mn}}$ by traversing the image in four directions, top, bottom, left and right. During the traversal, the white pixel areas are removed in all directions. Finally we get the textual area image $\overline{\overline{G(d_i)_{mn}}}$ which having the textual part of the document image $(d_i)_{mn}$. The textual image $\overline{\overline{G(d_i)_{mn}}}$ is mapped to the original image $(d_i)_{mn}$ and the resultant image from the preprocessing process is denoted as,

$$P(d_i)_{mn}; m = 1,2,\cdots r, n = 1,2,\cdots s \quad (2)$$

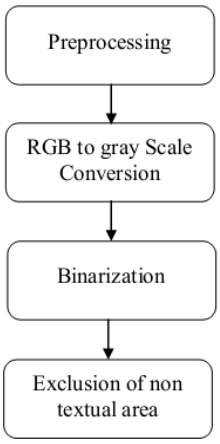


Figure 2: Procedure Of Preprocessing

3.2 Adaptive Histogram Equalization

Histogram equalization frequently raises the local contrast of numerous images, especially when the functional data of the image is signified by close difference values. The handwritten image documents texts to be blurred and we can't make out the text lines more precisely. Thus we raise the difference of the document image texts to acquire the precise segmentation result. We work out the adaptive histogram equalization technique for improving the image documents texts in our suggested method.

Initially, the adaptive histogram equalization is calculated on the preprocessed image $P(d_i)_{mn}$. The histogram Equalization is stated as follows,

Journal of Theoretical and Applied Information Technology
20th December 2013. Vol. 58 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$h_{mn} = \sum_{m=1}^{r}\sum_{n=1}^{s} P(d_i)_{mn} \quad (3)$$

Following the histogram equalization which is given in Equ. (3), the binarization method is executed and the binary image result is represented as h_{mn}'. After that, the mean histogram and threshold values are calculated for accomplishing the line segmentation process.

The mean histogram is calculated by,

$$\mu_{mn} = \sum_{m=1}^{r}\sum_{n=1}^{s} h_{mn}' \quad (4)$$

Subsequently the optimal threshold value is found for sorting out the textual positions and non textual locations so as to fragment each line in the document image. A most favorable threshold value that divides the two lines is evaluated as below.

$$t = \frac{1}{r}\sum_{m=1}^{r}\sum_{n=1}^{s} \mu_{mn} \quad (5)$$

The overall process after the histogram equalization is illustrated in Figure. 3.

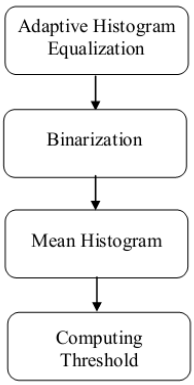


Figure 3: Process Of Threshold Computation

3.3 Sliding Window Based Line Segmentation

Fragmenting the document images encompassing skewed lines and lines with overlapping manuscript by application of the uncomplicated, threshold based projection techniques will not offer precise segmentation results since these techniques are very responsive to still a single noise. Moreover, the go beyond characters in the document may raise the projection values of non-textual place beyond the threshold value which deceives to an inappropriate segmentation. A sliding window method is applied in our proposed approach to address this problem. Usually the handwritten texts have line skew in the upper or lower side of the base line. To fragment the skewed line, the three sliding windows containing the height equal to the optimum threshold value and estimated to variant angles are traversed along the histogram of the document image. Amongst the three sliding windows, one is projected along the base line, the another window is projected to upper side of the base line and the left behind one is projected to lower side of the base line. The three sliding windows $s_1(\theta_1)$, $s_2(\theta_2)$ and $s_3(\theta_3)$ are made and crossed horizontally beside the 'μ_{mn}' in step by step manner till the value of's'. While crossing, for each movement, for the three window element, the histogram values of the image engaged by the window areas

$$w_k(s_i(\theta_i)) \quad (6)$$

In Equ. (6), where 'i' is the no of windows (i=3) and 'k' is the number of movements. For every movement, the area of the window containing high $h(w_k(s_i(\theta_i)))$ is regarded for fragmentation. Following a single straight traversal, all the $h(w_k(s_i(\theta_i)))$ attained are cumulative and capitulated the first line of the text, next the window position is shifted down and the then horizontal window crossing is initiated from 1 to 's' as above, the subsequent line is fragmented. Therefore three windows are crossed from left to right for the complete image and all the lines of the handwritten text images are fragmented.

3.4 Skewing Operation

The ensuing line fragmented text image $P(d_i)_{mn}''$ is given to the Skewing process. During text scanning or copying, document skew is a deformation that mostly happens. Document skew is an obvious result because of the composite arrangement of handwritten words and the copying/scanning process, especially when digitizing automatically huge text bulks. The twisted images require to be de-skewed for precise fragmentation. At first the allied elements are recognized from the morphological enlarged image and next the optimum skew angle of the joined component is found out by discovering centroid of every connected element in the text and designing ellipse on it. The optimum angle is found by the rotation of the input image. The skewed line fragmented text result image is $R(d_i)_{mn}''$.

4. EXPERIMENTAL RESULTS AND DISCUSSION

The proposed handwritten text line segmentation technique for Tamil language based on sliding

Journal of Theoretical and Applied Information Technology
20th December 2013. Vol. 58 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

window and skewing is implemented in the working platform of MATLAB R2011a (Version 7.12) with machine configuration as given in Table 1.

TABLE 1: Machine configuration

Processor	Intel core i5
OS	Windows XP
CPU Speed	3.20GHz
RAM	4GB

The sample handwritten text document images are shown in Figure. 4.

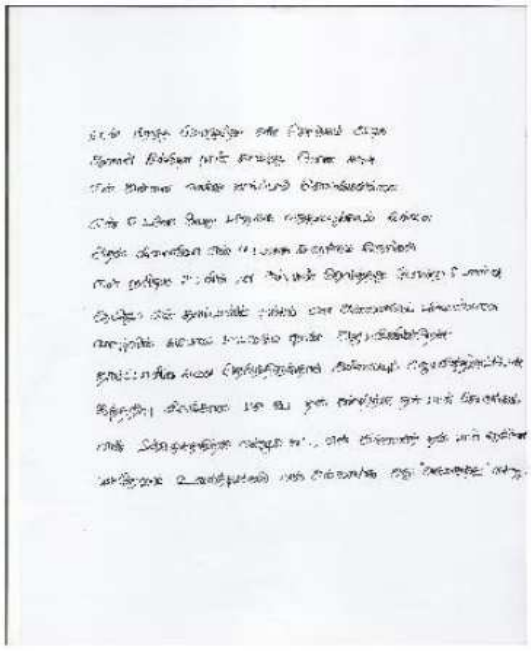
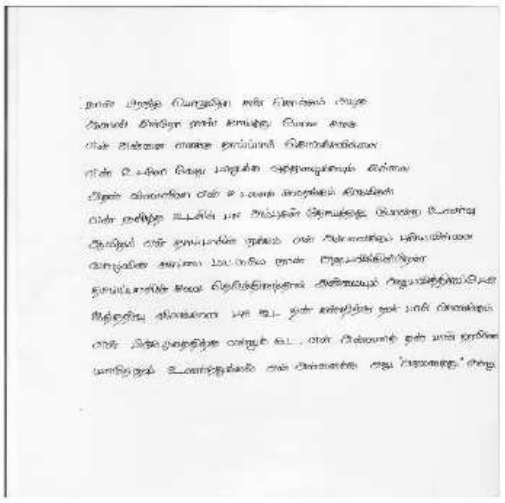
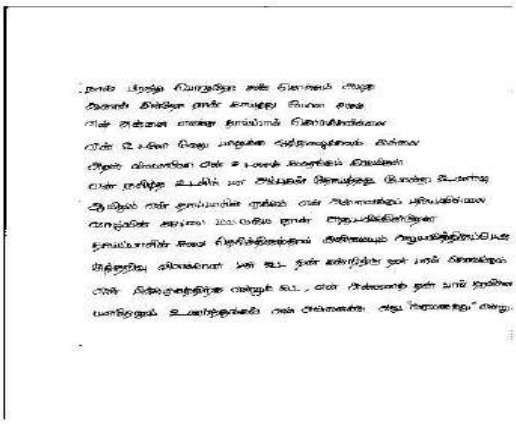


Figure 4: Sample Tamil Handwritten Text Document Image

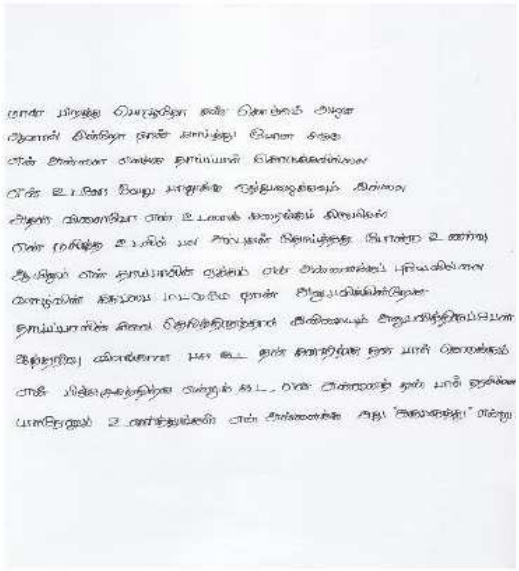
The sample handwritten document image were given to the preprocessing process, the sample document result from the preprocessing results are given in Figure. 5.



(i)



(ii)



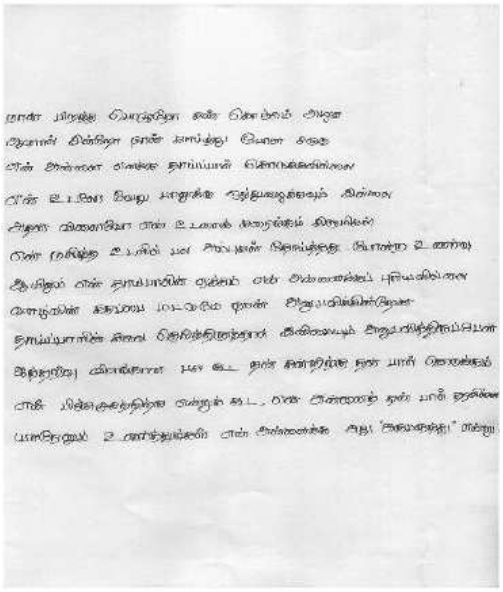
(iii)

Figure 5: Preprocessing Result (I) RGB To Grayscale (Ii) Binarization And (Iii) Removal Of Non Textual Area

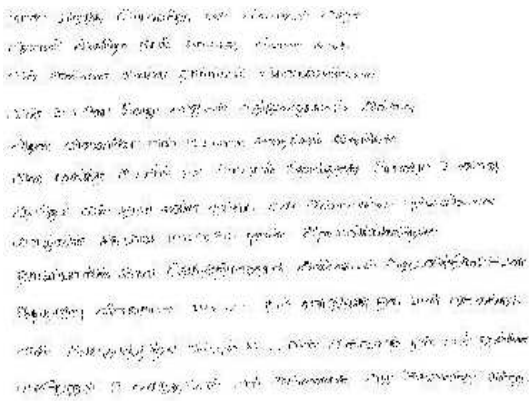
Journal of Theoretical and Applied Information Technology
20th December 2013. Vol. 58 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The adaptive histogram equalization and the line segmentation using the sliding window and skewing operations results are shown in Figure. 6 and 7.



(i)



(ii)

Figure 6: (i) Adaptive Histogram Equalized image and (ii) Binarization of Adaptive Histogram Equalized image

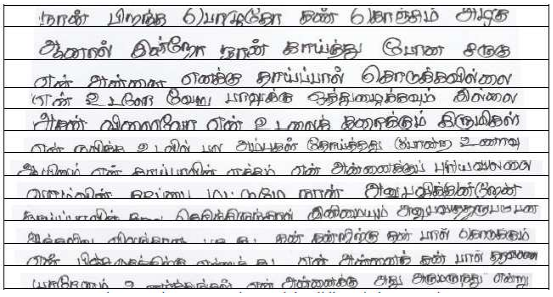


Figure 7: Line Segmentation Result By Sliding Window Operation

4.1 Performance and Comparative Analysis

The performance of the proposed Tamil language handwritten text line segmentation technique is evaluated with 20 handwritten text document images. These 20 text document images are given to the segmentation process and this proposed technique results are compared with the conventional segmentation technique [26]. For performance analysis the documents dataset is divided into two datasets namely Dataset 1 and Dataset 2, and each dataset contains 10 handwritten text document images. The proposed and the conventional techniques performance are evaluated by the three performance measures namely DR (Detection Rate), RA (Recognition Accuracy) and FM which are given in [27]. These performance measures results for the proposed and conventional techniques are given in Table 2. Those measures comparison graph between the proposed and the conventional techniques is shown in Figure. 8.

TABLE 2: Performance Of Proposed And Conventional Techniques (I) Dataset 1 And (Ii) Dataset 2

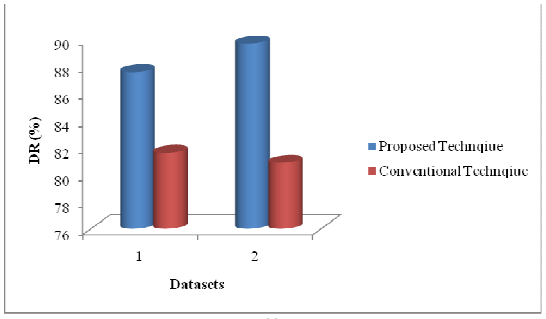
Number of Text Documents	Proposed Technique			Conventional Technique		
	DR (%)	RA (%)	FM (%)	DR (%)	RA (%)	FM (%)
1	79.23	70.32	74.51	73.13	64.02	68.27
2	80.32	82.93	81.60	78.02	80.73	79.35
3	84.34	81.02	82.65	73.34	76.02	74.66
4	82.54	87.49	84.94	72.54	77.09	74.75
5	86.57	79.41	82.84	66.41	68.11	67.25
6	93.21	92.98	93.09	73.32	72.98	73.15
7	92.63	90.67	91.64	72.36	70.77	71.56
8	95.21	94.84	95.02	85.01	83.54	84.27
9	90.87	89.47	90.16	88.36	86.37	87.35
10	90.67	91.63	91.15	85.43	87.33	86.37

(i)

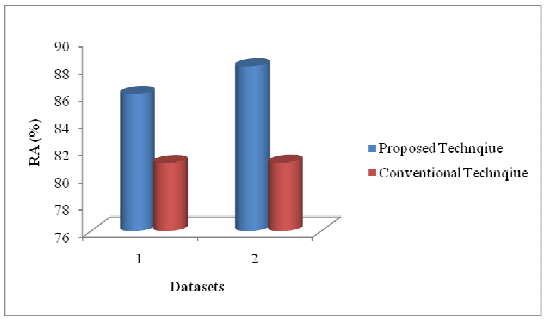
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Number of Text Documents	Proposed Technique			Conventional Technique		
	DR (%)	RA (%)	FM (%)	DR (%)	RA (%)	FM (%)
1	80.32	75.42	77.79	73.22	76.32	74.74
2	86.34	88.39	87.35	76.24	78.59	77.40
3	88.53	83.20	85.78	78.33	80. 07	79.19
4	85.53	84.89	85.21	75.83	74.49	75.15
5	85.65	83.14	84.38	80.65	81.14	80.89
6	93.27	95.48	94.36	73.07	75. 18	74.11
7	94.33	94.67	94.50	84.33	84.27	84.30
8	95.21	94.84	95.02	75.11	74.44	74.77
9	94.87	89.47	92.09	84.47	85.47	84.97
10	92.67	91.63	92.15	72.37	71.43	71.90

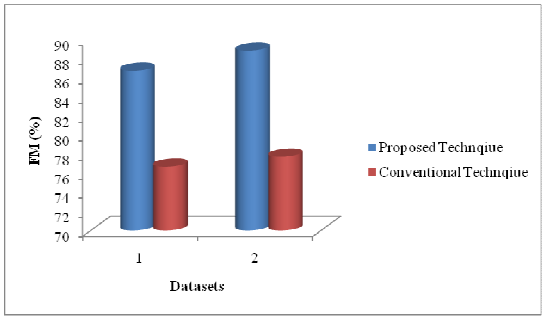
(ii)



(i)



(ii)



(iii)

Figure 8: Averaged Comparison Graph Of Proposed And Conventional Techniques In Terms Of (I) DR (Ii) RA And (Iii) FM

Discussion: As shown in Figure. 8 and Table 2, the experimental results indicate that the proposed technique has higher segmentation result than the conventional technique. Our proposed segmentation technique has achieved averaged 88%, 87% and 88% of DR, RA and FM respectively. Compared to the conventional segmentation technique, our proposed technique has given averaged 6%, 7% and 10% higher DR, RA and FM respectively. Thus the performance metrics results shows that our proposed text line segmentation technique for Tamil language has more efficiently segments the text lines than the conventional technique.

5. CONCLUSION

In this paper, we proposed an efficient Tamil language handwritten text line segmentation technique which segments the input handwritten text lines more accurately. Initially, the input handwritten documents binarization was computed by the preprocessing and these preprocessed images were given to the adaptive histogram equalization. By using this histogram equalized images, the text lines were accurately segmented by exploiting sliding window and skewing operations. The performance of the proposed Tamil language handwritten text line segmentation technique was evaluated by the more number of text line documents. The implementation result shows that the proposed Tamil language handwritten text line segmentation technique efficiently segments the text lines than the conventional technique. Experimental results show that the proposed Tamil language handwritten text line segmentation technique has attains high DR, RA and F-Measure value than conventional technique.

REFERENCES

[1] A. Nicolaou and B. Gatos,"Handwritten Text Line Segmentation by Shredding Text into its Lines", *In preceding of 10th International Conference on Document Analysis and Recognition*, pp. 626-630, 2009.

[2] David Doermann, Jayant Kumar, WaelAbd-Almageed and Le Kang, "Handwritten Arabic

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

text line segmentation using affinity propagation", *In proceedings of the 9th IAPR International Workshop on Document Analysis Systems*, pp. 135-142, 2010.

[3] Yi Li, YefengZheng, David Doermann and Stefan Jaeger "A New Algorithm for Detecting Text Line in Handwritten Documents", *In proceedings of 10th International Workshop on Frontiers in Handwriting Recognition, La Baule*, France,pp. 35-40, 2006.

[4] G. Louloudis, B. Gatos, I. Pratikakis and C. Halatsis,"Text line and word segmentation of handwritten documents", *Pattern Recognition*, Vol. 42, No. 12, pp. 3169–3183, December 2009.

[5] Â RejeanPlamondon and Sargur N. Srihari, "On-Line and Off-Line Handwriting Recognition: A Comprehensive Survey", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 1, pp. 63-84, January 2000.

[6] VassilisPapavassiliou, ThemosStafylakis, VassilisKatsourosand George Carayannis, "Handwritten document image segmentation into text lines and words", *Pattern Recognition* Vol. 43,pp. 369-377, 2010.

[7] Fei Yin and Cheng-Lin Liu,"Handwritten Chinese text line segmentation by clustering with distance metric learning", *Pattern Recognition*, Vol. 42,pp. 3146-3157, 2009.

[8] Mahdi KeshavarzBahaghighat and JavadMohammadi, "Novel Approach for Baseline Detection and Text Line Segmentation", *International Journal of Computer Applications*, Vol. 51, No.2, pp. 9-16, August 2012.

[9] ZaidiRazak, KhansaZulkiflee, Mohd Yamani IdnaIdris, EmranMohd Tamil, MohdNoorzaily, Mohamed Noor, RosliSalleh, MohdYaakob, ZulkifliMohdYusof, and MashkuriYaacob,"Off-line Handwriting Text Line Segmentation: A Review", *International Journal of Computer Science and Network Security*, Vol.8 No.7, pp. 12-20, July 2008.

[10] J. Venkatesh and C. Suresh kumar, "Tamil Handwritten Character Recognition Using Kohonon's Self Organizing Map", *International Journal of Computer Science and Network Security*, Vol.9 No.12, pp. 157-161, December 2009.

[11] Chen Huang andSargur N. Srihari,"Word segmentation of off-line handwritten documents", *In Proc. SPIE 6815, Document Recognition and Retrieval XV*, Vol. 6815, No.1, pp. 1-5, 2008.

[12] N. Sengottaiyan and C.Sureshkumar"Handwritten South Indian Language Recognition Using Artificial Neural Network", *International Journal of Advanced Research in Technology* Vol. 1, No. 1, pp. 87-90, Sep 2011.

[13] C.SureshkumarandT.Ravichandran, "Handwritten Tamil Character Recognition and Conversion using Neural Network", *International Journal on Computer Science and Engineering*, Vol. 02, No. 07, 2261-2267, 2010.

[14] R. Manmatha and NitinSrimal, "Scale Space Technique for Word Segmentation in Handwritten Documents", *Scale-Space Theories in Computer Vision*, Vol. 1682, pp. 22-33, 1999.

[15] MamathaHosalliRamappa and Srikantamurthy Krishnamurthy, "Skew Detection, Correction and Segmentation of Handwritten Kannada Document", *International Journal of Advanced Science and Technology*, Vol. 48, pp. 71-88, November, 2012.

[16] Richard G. Casey and Eric Lecolinet, "A Survey of Methods and Strategies in Character Segmentation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 18, No. 7, pp. 690-706, 1996.

[17] P Shivakumara, G Hemantha Kumar, D S Guru And P Nagabhushan"A novel technique for estimation of skew in binary text document images based on linear regression analysis", *Sadhana*, Vol. 30, No. 1, pp. 69–85, February 2005.

[18] S.V. Rajashekararadhya and P. VanajaRanjan,"Efficient Zone Based Feature Extration Algorithm for Handwritten Numeral Recognition of Four Popular South Indian Scripts", *Journal of Theoretical and Applied Information Technology*,

[19] Basappa B. Kodada and Shivakumar K. M. "Unconstrained Handwritten Kannada Numeral Recognition", *International Journal of Information and Electronics Engineering*, Vol. 3, No. 2, pp. 230-232, March 2013.

[20] Mamatha H.R, Srikanta Murthy K, Sudan S, Vinay G Raj and Sumukh S Jois,"Fan Beam Projection Based Features to Recognize Handwritten Kannada Numerals*", In preceding of International Conference on Software and Computer Applications, Singapore*, Vol. 9, pp. 257-262, 2011.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[21] Apurva A. Desai, "Gujarati handwritten numeral optical character reorganization through neural network", *Pattern Recognition*, Vol. 43,pp. 2582–2589, 2010.

[22] YousriKessentini, Thierry Paquet and AbdelMajid Ben Hamadou, "Off-line handwritten word recognition using multi-stream hidden Markov models", *Pattern Recognition Letters*, Vol. 31,pp. 60–70, 2010.

[23] V.N. ManjunathAradhya and G. Hemantha Kumar, S. Noushath "Multilingual OCR system for South Indian scripts and English documents: An approach based on Fourier transform and principal component analysis", *Engineering Applications of Artificial Intelligence*, Vol. 21, pp. 658–668, 2008.

[24] MamathaHosalliRamappa, and Srikantamurthy Krishnamurthy, "Skew Detection, Correction and Segmentation of Handwritten Kannada Document", *International Journal of Advanced Science and Technology*, Vol. 48, pp. 71-88, 2012.

[25] Mamatha H R and Srikantamurthy K, "Morphological Operations and Projection Profiles based Segmentation of Handwritten Kannada Document", *International Journal of Applied Information Systems*, Vol. 4, No.5, pp. 13-19, October 2012.

[26] Bhowmik, A. Roy and U. Roy, "Character Segmentation for Handwritten Bangla Words Using Artificial Neural Network", *In Proceedings of the International Workshop on Neural Networks and Learning in Document Analysis and Recognition*, 2005

[27] http://en.wikipedia.org/wiki/Precision_and_recall