Special Issue | August 2017 ISSN: 2320-9984 (Online)

International Journal of Modern Engineering & Management Research

Website: www.ijmemr.org

All India Seminar on

Futuristic Trends in Telecommunication Engineering & Telecom Panorama – Fundamentals and Evolving Technology, with Particular Reference to Smart City on 5th – 6th August 2017

Organized by

The Institution of Engineers (India) Jabalpur Local Centre

Text Line Segmentation Based on Horizontal Projection Method

Ashna Joysi Gaikwad
M. Tech, (Digital Communication)
Gyan Ganga College of Technology
Jabalpur (M.P.) INDIA
Email: ashnajoysi@gmail.com

Siddharth Bhalerao
Assistant Professor
Department of Electronics and Communication
Gyan Ganga College of Technology
Jabalpur (M.P.) INDIA
Email: siddharthbhalerao@ggct.co.in

Abstract**—***Optical Character Recognition is a process which converts the handwritten or printed text in a machine readable form. The segmentation is a part of optical character recognition (OCR) which derives text from the non text region in a document. The text is segmented into lines then words then characters. In this paper we will deal with Line segmentation using Horizontal Projection Profile. This paper investigates the proposed algorithm with Hindi, English and text with skewed lines. The accuracy of segmentation of these datasets were compared.*

Keywords:—*Optical character recognition, Segmentation, Horizontal Projection Profile, handwritten texts.*

1. INTRODUCTION

Optical character recognition is a part of image processing that allows machine to automatically recognize characters through a mechanism. Text segmentation is an important task in any OCR. Deriving text from non text region in a document is called segmentation. The text line segmentation is the initial step of character segmentation. The text line segmentation can be done through various ways.

A lot of research has been done in line segmentation existing methods like Projection Profile method [1], Cut Text Minimum method [2], Smearing method [3], liner Programming method [4].and Hough transform method [5].

The objective of this paper is to use a method based on Projection profile to segment text in Hindi, English and text with skewed lines then concluding the overall performance of this method. This paper is organized as follows: Section 2 describes the complexity of a text document, Section 3 explains the proposed algorithm and section 4 contains Result and section 5 contains Discussion.

2. COMPLEXITY OF A TEXT DOCUMENT

Any line is divided into three regions i.e. upper zone, middle zone and lower zone. Apart from this a line may have overlapping and touching components [6]. These make line segmentation process complex. The author style and line orientation may give different results in segmented lines. So for an accurate outcome standard text consisting of uniform lines and proper orientation should be used.

3. PROPOSED METHOD

The method used comes under Horizontal Profile Projection which shows the sum of ON pixels along a row. The ON pixels are the text lines and the OFF pixels depict the space between the lines. The algorithm is given as follows:

Begin

Input: Handwritten Document

Output: Segmented lines

Step 1: The handwritten text document is converted into binarised text using OTSU threshold [7].

Step 2: Morphological operations are applied. Erosion followed by dilation [8].

Step 3: The horizontal projection profile (HPP) of the image is obtained.

Step 4: Observing the HPP, centroids of the blank regions are found i.e. region between the peaks.

Step 4: Using the centroids of blank regions as the delimiters lines are segmented from the text.

End

4. RESULTS

For dataset we have used three handwritten documents in Hindi, English and text with skewed lines. The dataset is taken in .JPG format. For the diversity in dataset they have been written with different orientation and pen tips. The dataset include following:

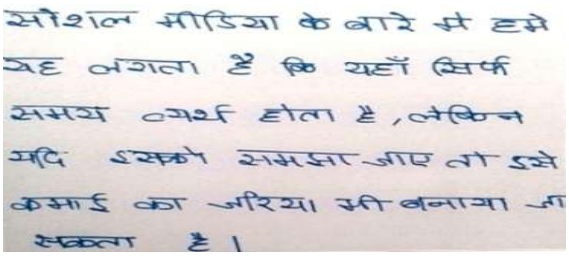

Figure 1. Handwritten Hindi text

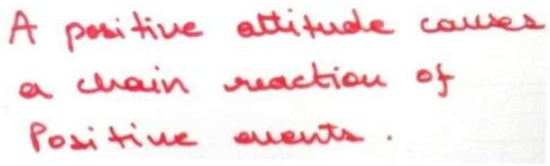

Figure 2. Handwritten English Text

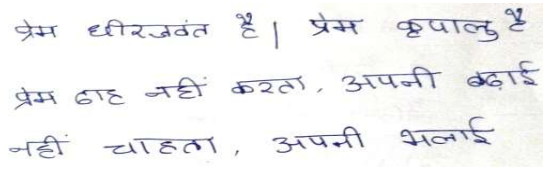

Figure 3. Text with Skewed Lines

After using the proposed method the following outputs have been obtained:

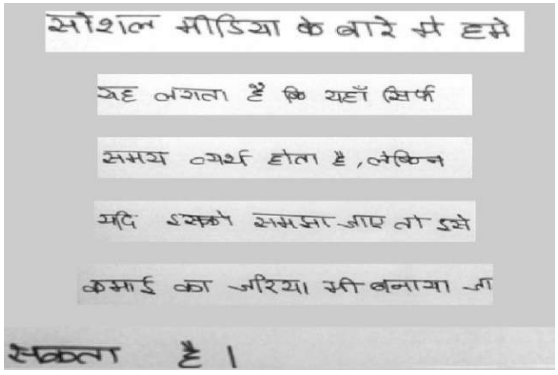

Figure 4. Segmented lines of figure 1

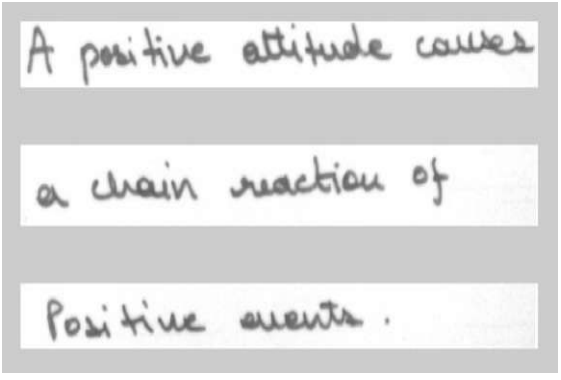

Figure 5. Segmented Lines of figure 2

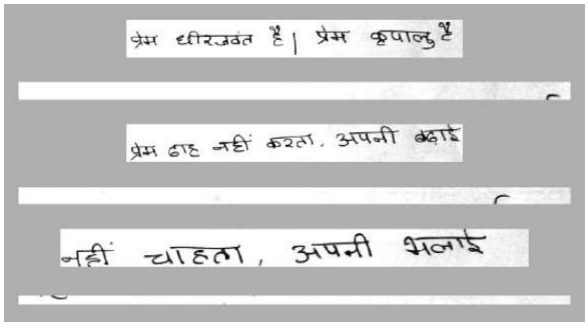

Figure 6. Segmented lines of figure 3

The accuracy is calculated by :

$$Accuracy\% = \frac{Lines\ Correctly\ Segmented}{Total\ number\ of\ lines} \times 100$$

Comparative study of different datasets with proposed algorithm is given in Table I.

Table 1. Accuracy of Text Line Segmentation

Text	Total number of lines	Lines correctly segmented	Accuracy
Hindi Text	6	6	99.9%
English Text	3	3	99.9%
Text with Skewed Lines	3	2.5	83.4%

5. DISCUSSIONS

From the result we observe that the proposed method is independent of the script of document. This method gave accurate result with text with straight lines in Hindi and English.

But with the case of skewed lines we find that the lines are segmented correctly at the cost of ascenders and decenders.

This can be seen in  which is segmented in two lines i.e. the consonant part with rest of the lines as  and the curve part of *matra* is segmented in the line above it as  .

Hence it is estimated that this method may give good results with skewed lines if its allographs get smaller than upper zone of the line.

The study may be carried out in future with following directions:

Segmentation method can be improved to segment skewed lines.

The segmentation can be applied with the characters also.

This segmentation can be applied to other Indian scripts.

REFRENCES:

[1] Chethana H T, Mamatha H R, *Comparative Study of Text Line on Handwritten Kannada Documents, (IJCSIT) International Journal of Computer* Science and Information Technologies, Vol. 7 (1), 26-33,ISSN 0975-9646, 2016

[2] Saiprakash Palakollu, Renu Dhir and Rajneesh Rani, *A New Technique for Line Segmentation of Handwritten Hindi Text,* (ICEICE), Special Issue of International Journal of Computer Applications on Electronics, Information and Communication Engineering - (0975 – 8887) ICEICE No.5, Dec 2011

[3] Chethana H T, Mamatha H R, *Comparative Study of Text Line on Handwritten Kannada Documents, (IJCSIT) International Journal of Computer* Science and Information Technologies, Vol. 7 (1), 26-33,ISSN 0975-9646, 2016

[4] Zaidi Razak, Khasana Zulkilfe and Mashkuri Jacob, *OffLine Handwriting Text Line Segmentation: A Review,*

(IJCSNS), International Journal of Computer Science ans Network Security,Vol.8, No. 7,July 2008

[5] G. Louloudis, B. Gatos, I. Pratikakis and K. Halatsis, "*A Block Based Hough Transform Mapping for Text Line Detection in Handwritten Documents*", in the proceedings of Tenth International Workshop on Frontiers in Handwriting Recognition, La Baule, pp. 515-520, 2006.

[6] Jatinder Kaur, Er. Navdeep Singh Sethi, *Segmentation of Broken and Touching Characters in Handwritten Gurumukhi Word using Clustering Approach*, Volume 5, Issue 7, July 2015, ISSN: 2277 128X

[7] Puneet GRDIET, Naresh Kumar Garg, Binarization Techniques used for Grey Scale Images, International Journal of Computer Applications (0975 – 8887) Volume 71– No.1, June 2013

[8] Mamatha HR and Srikantamurthy K, Morphological Operations and Projection Profiles based Segmentation of Handwritten Kannada Document, International Journal of Applied Information Systems (IJAIS)–ISSN:2249-0868 Foundation of Computer Science FCS, 2012.