IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH - TECHNOLOGY

ADVANCED PROFILE PROJECTION TECHNIQUE FOR CHARACTER SEGMENTATION OF HANDWRITTEN TEXT IN GURUMUKHI SCRIPT

Ankita Chanana*, Chandana Jain
*Student of computer science and engeenring, Jan Nayak Chaudhary devilal memorial college, Sirsa, Haryana, India
Assistant professor of cse department, Jan Nayak Chaudhary devilal memorial college, Sirsa, Haryana, India

ABSTRACT
Character segmentation is current topic of discussion for today's research work. Because of its importance in recognition of characters. It is the most basic step in the process of recognition of characters But it is not an easy job to separate characters from words specially in handwritten text because writing styles vary from person to person. Not only this there are other so many problems associated with it like overlapping of characters, touching of characters, broken characters, skewed characters. Besides that there is also the concept of upper zone, lower zone and middle zone. This paper presents an Advanced profile projection technique for character segmentation which deals with overlapping of characters, touching characters, broken characters in handwritten text o Gurumukhi script.

KEYWORDS: Character Segmentation, Profile projection technique

INTRODUCTION
Gurumukhi script is used to write Punjabi language. The writing style of Gurumukhi script symbol is from left to right side of paper. In Gurumukhi script there is no upper or lower case character concept. There are 41 consonants and 12 vowels in the Gurumukhi script.

In optical character recognition(OCR), segmentation is a significant phase and accuracy of recognition highly depends on accuracy of segmentation. Incorrect segmentation leads to incorrect recognition. OCR is a useful invention which is used to read and recognise the scanned documents. The main aim of OCR is to convert the scanned document into readable or editable format. Steps in OCR-

optical scanning
preprocessing
#segmentation
#feature extraction
#recognition
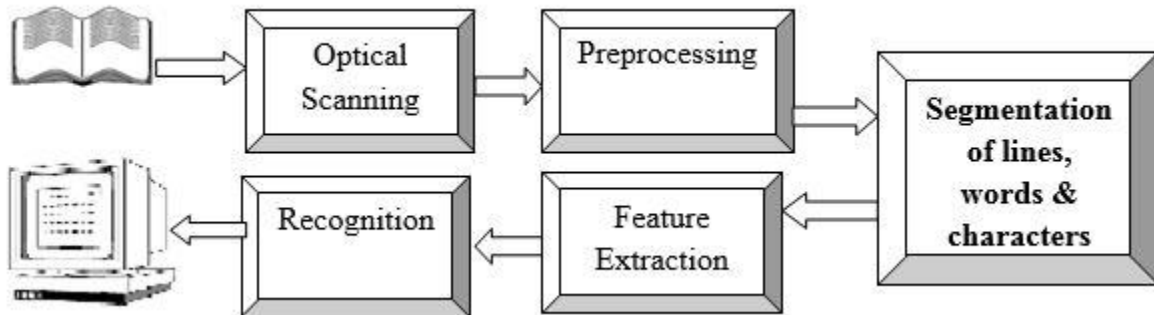
Character segmentation is the technique used to separate charachtres from one character to another. There are various algorithms to segment the Gurumukhi handwritten words into characters. But there can be overlapped characters, touching characters and also characters with upper and lower zones due to which segmentation becomes difficult. So

in this paper we tried to segment the characters og all the three zones and also removed problems like overlapping, touching and broken characters. We have developed an algorithm that will :-
Segment the isolated character
Identify the presence of touching character
Find out the break point
Segments the touched character
Also separates all the three zones

LITERATURE SURVEY
In today's Research Senarios, there are different techniques which have been discussed for character segmentation. Segmentation on handwritten text is difficult process then segmentation of characters on printed document.
An end detection algorithm of segmentation of broken and touching characters in Handwritten Gurumukhi word" A survey done by Parika Mangla and Harleen Kaur(**IEEE**)**.** The paper provides a new segmentation technique based on neighbouring pixels for touching and broken characters of Handwritten Punjabi text that is the Gurumukhi Script.[1].
"Fragmentation of Handwritten Touching Characters in Devnagari Script" A survey by Shuchi Kapoor and Vivek Verma. In which a technique is developed to provide a solution to the touching characters[2]
"To Extract Feature of Handwritten Devnagari Script" A feature Extraction technique is used by Ajay Garg and Simpel Jindal to recognise handwritten Devnagari Script document[3].
"Text Line Detection and Segmentation in Handwritten Gurumukhi Scripts" An effective method is proposed by Namisha Modi and Khushneet jindal for text line segmentation in Handwritten Punjabi document that deals with the problems like overlapped and connected components[4].
"Handwritten Hindi Text Segmentation Techniques for Line and Characters" a survey done by Saiprakash palakollu, Renu Dhir and Rajneesh Rani. This paper deals with the various methods of line and character segmentation.. In this paper the basic technique which is followed is that header lines are detected and converted as straight lines. After that, each word is divided into upper modifier then consonant and then into lower part, to make character segmentation easy. Algorithm is based on finding header and base lines by estimating the average line height. This technique efficiently segments lines with accuracy upto 93%, segment words with accuracy upto 96% and characters with accuracy upto 89%. This method for line segment is working efficiently in the cases of different text sizes and different resolution.[5]

"The Hazards in segmentation of Handwritten Hindi Text" In this paper Naresh Kumar Garg, Lakhwinder Kaur and M.K. Jindal done a very good job, This paper provides an over all view of the problems which are currently exist in Handwritten text[6].
"Segmentation of Handwritten Hindi Text" this survey is also done by Naresh Kumar Garg, Lakhwinder Kaur and M.K. Jindal also work on Segmentation of Handwritten Hindi Text. In this paper new segmentation technique based on structural approach is provided [7].
"Isolated Handwritten Words Segmentation Techniques in Gurumukhi Script" In this paper Galaxy Bansal and Dharamveer Sharma worked on Segmentation of Isolated words in Gurumukhi Script. The main objective of this paper was to discuss a compele solution for character segmentation phase of Gurumukhi Script[8].
touching characters[2]
"To Extract Feature of Handwritten Devnagari Script" A feature Extraction technique is used by Ajay Garg and Simpel Jindal to recognise handwritten Devnagari Script document[3].
"Text Line Detection and Segmentation in Handwritten Gurumukhi Scripts" An effective method is proposed by Namisha Modi and Khushneet jindal for text line segmentation in Handwritten Punjabi document that deals with the problems like overlapped and connected components[4].
"Handwritten Hindi Text Segmentation Techniques for Line and Characters" a survey done by Saiprakash palakollu, Renu Dhir and Rajneesh Rani. This paper deals with the various methods of line and character segmentation.. In this paper the basic technique which is followed is that header lines are detected and converted as straight lines. After that, each word is divided into upper modifier then consonant and then into lower part, to make character segmentation

easy. Algorithm is based on finding header and base lines by estimating the average line height. This technique efficiently segments lines with accuracy upto 93%, segment words with accuracy upto 96% and characters with accuracy upto 89%. This method for line segment is working efficiently in the cases of different text sizes and different resolution.[5]

"The Hazards in segmentation of Handwritten Hindi Text" In this paper Naresh Kumar Garg, Lakhwinder Kaur and M.K. Jindal done a very good job, This paper provides an over all view of the problems which are currently exist in Handwritten text[6].

 "Segmentation of Handwritten Hindi Text" this survey is also done by Naresh Kumar Garg, Lakhwinder Kaur and M.K. Jindal also work on Segmentation of Handwritten Hindi Text. In this paper new segmentation technique based on structural approach is provided [7].

"Isolated Handwritten Words Segmentation Techniques in Gurumukhi Script" In this paper Galaxy Bansal and Dharamveer Sharma worked on Segmentation of Isolated words in Gurumukhi Script. The main objective of this paper was to discuss a compele solution for character segmentation phase of Gurumukhi Script[8].

"Segmentation of Printed Text in Devanagri Script and Gurumukhi Script" A survey is done by vijay kumar and Pankaj K. Sengar on printed text in Devanagri and Gurumukhi Script. This paper deals with the line, word, character and top character segmentation for printed Hindi text in Devanagari Script and it it also describe the line and word segmentation for printed text in Gurumukhi Script. In this paper a single algorithm is proposed for segmentation of Devnagari Script and Gurmukhi Script. A performance of 100% at line level, approximately 100% at word level , 99% at character level and 97% top character level.[9]

 "A Study of Different kinds of Degradation in printed Gurumukhi Script" M.K. Jindal, R.K. Sharma and G.S. Lehal studied different kinds of degradation in printed text and provide solution to some of them.[10].

PROPOSED WORK

In the proposed work we implement an algorithm along with various techniques along with horizontal profile projection technique, vertical profile projection technique to segment the characters of all the three zones. Single algorithm has been developed to segment the isolated characters, broken characters, touching characters of middle zone And it also segments upper zone and lower zone from a handwritten word written in Gurumukhi script. To implement this algorithm we have performed the following various steps:-

1.   Scanning image :-
In this step the document is converted into scanned image with the help of scanner. Set the threshold value of image upto 200

2.   Binarisation
In this step gray scale image image can be converted into binary image with the help of O.C.R software. Binarization is the process in which grey scale images are converted into 0's and 1's

3.   Calculate the width and height of the word:-
After the process of binarization the width and height of the word is calculated by calculating the number of pixels.

4.   Number of characters

5.   After calculating the height and width we can calculate the number of characters present in the word.

6.   Calculate the Header and bottom line;-
There are two main lines in which the word is placed. The upper most line is called the header line and the bottom most line is called the bottom line.

7.   Removal of header line:-
the next important step after binarization of image is the removal of header line. In this step the header line which puts all the characters together in a word is removed.
   Steps:
          #Calculate the frequency of black poxels in each row

Row having maximum number of pixels is considered as the headel line
Replace all the 1's in that row with o's

7 Segment the character

Find the gap between the characters of a given word. Then find the mid of gap. Segment the character from this mid value
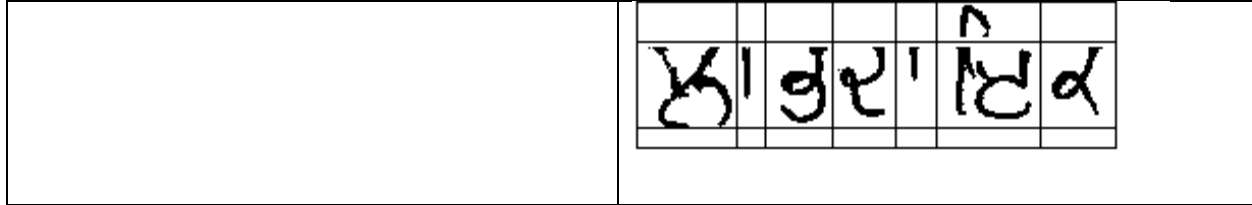
8 Touching characters:-

If gap between the mid point is large then 2*size, then it means that characters are touching.find the total number of touching characters. Find the end of these characters. Segment the characters from these points

9 Repeat the code for all the characters.

10 End

Word	Input image	Output image
Isolated		
Broken		
Touching		

Input image(complicated)	Output image (with zoning)

CONCLUSION AND FUTURE WORK

Here we have tested this algorithm on 150 handwritten taken from different people. In which there was isolated, touching and broken words and also it performs zoning of characters that is character is segmented into different zones. word in handwritten document can consists of many simple characters, touching characters, broken characters, skewed characters. We tried to remove so many problems but some are still left like skewed characters cannot be segmented properly. So this system can be extended to remove the remaining problems which exsist in handwritten documents.

REFERENCES

1. Parika Mangla,Harleen Kaur"An end detection algorithm of segmentation of broken and touching characters in Handwritten Gurumukhi word" in IEEE,2014.
2. Shuchi Kapoor and Vivek Verma "Fragmentation of Handwritten Touching Characters in Devnagari Script"in IJITMC,2014 .
3. Ajay Garg and Simpel Jindal "To Extract Feature of Handwritten Devnagari Script"in IJ of Advanced Research in Computer and Communication,2014.
4. Namisha Modi and Khushneet Jindal "Text Line Detection and Segmentation in Handwritten Gurumukhi Scripts"in IJ of Advanced Research in Computer Science and Software Engeenring, 2013.
5. Saiprakash palakollu, Renu Dhir and Rajneesh Rani "Handwritten Hindi Text Segmentation Techniques for Line and Characters" in WCECS 2012.
6. Naresh Kumar Garg, Lakhwinder Kaur and M.K. Jindal "The Hazards in segmentation of Handwritten Hindi Text" in International Journal of Computer Applications,2011
7. Naresh Kumar Garg, Lakhwinder Kaur and M.K. Jindal "Segmentation of Handwritten Hindi Text" in International Journal of Computer Applications,2010
8. Galaxy bansal, Dharamveer Sharma "Isolated Handwritten Words Segmentation Techniques in Gurumukhi Script"in IJ of computer applications,2010
9. Vijay kumar and Pankaj K. Senagar "Segmentation of Printed Text in Devanagri Script and Gurumukhi Script" in IJCA 2010.
10. M.K. Jindal, R.K. Sharma and G.S. Lehal "A Study of Different kinds of Degradation in printed Gurumukhi Script" in IEEE, 2007.