International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 02 | Feb 2021 www.irjet.net p-ISSN: 2395-0072

Identification of Offline Gujarati Handwritten Conjunct Characters

Mayuri Patel

Dept. of Computer, Institute of Hotel Management & Catering Technology Silvassa, Dadra & Nagar Haveli, India

---

Abstract – *Optical character recognition is most popular technique in pattern recognition, which is used to recognize printed document into digital form. To recognize handwritten documents into machine readable form Handwritten Character Recognition is used, which is derived from Optical Character Recognition. Various scripts can be recognize by OCR, but the Devanagari script is very challenging task to recognize. A lot of work has been completed in Devanagari script recognition ,but still there is no more accuracy in Gujarati script especially for conjunct characters. To overcome conjuncted characters difficulties, we have to focus on characteristics of fusion characters. This paper presents a detailed review in the field of Gujarati Handwritten Conjunct Character Recognition.*

***Key Words*:**
Optical Character Recognition **(OCR)**, Handwritten Character Recognition **(HCR)**, Devanagari Script, Conjunct Characters, Digital Image Processing **(DIP).** Gujarati Handwritten Conjunct Characters **(GHCCR)**

1.INTRODUCTION

India is the country where various languages are used like Hindi, Gujarati, Tamil, Marathi, etc. to speak, listening and writing. In different sectors like Banking, Health Care, Education, Postal etc., are used handwritten documents. In current scenario our world is going to become digitizing. It is difficult and complex task to convert each and every old valuable documents into digital form and manually it is not feasible and also may not be secure. So, there are highly need to automation of old handwritten documents recognition. Recognition of Handwritten characters in different languages are sub domain of pattern recognition with digital image processing. Generally, to recognize characters from printed document or handwritten documents Optical Character. Recognition(OCR) is used. There are two types – Online Character Recognition & Offline Character Recognition. Optical character recognition extends Handwritten character recognition technique which is used to convert handwritten documents into machine readable/editable form[3].

1.1 Digital Image Processing

Image processing is a method to perform some operations on an image to get an enhanced image or to extract some useful information from an image. Digital Image Processing (DIP) is a type of signal processing in which input is an image and output may be image or features associated with that image.

1.2 Optical character recognition (OCR)

Optical character recognition is a process which converts printed or handwritten characters into digital form. Optical character recognition is one of the oldest and largest sub field of pattern recognition. Optical Character Recognition can improve a communication between human and machine in many applications. Now a day's various fields use OCR like Banking, Health Care, Industries etc . OCR is divided into two approaches , Online OCR and Offline OCR. An online OCR sense the movement of pen or other devices and converts it into digital form. In offline OCR input is scanned from images or printed documents and converted into digital form[1].

1.3 Handwritten Character Recognition(HCR)

Handwritten character recognition (HCR) is one of the part of image processing and pattern recognition. In HCR method, the input is scanned from images, documents and devices like computers, digitizers etc, which are converted into digital text form. Handwritten character recognition is done through two ways which are online handwritten character recognition and offline handwritten character recognition - Online Handwritten Character Recognition which takes the input at run time and Offline Handwritten Character Recognition which works on scanned images. Offline handwritten character recognition is one of the most challenging task in the field of image processing and pattern recognition in the recent years. There are so many languages can be recognize by HCR like Gujarati, Hindi, Marathi etc. Among these language Gujarati script is one of the most challenging task to recognize by HCR method .

2. Gujarati Handwritten Conjunct Character Recognition (GHCCR)

Gujarati is an Indo-Aryan language and one of the official languages of India, spoken by people of Gujarat state, union territories of Diu - Daman and Dadra - Nagar Haveli[1]. It is very difficult task to recognize Gujarati script because writing style of each person is differs from other person. A lot of work has been developed for Gujarati script recognition but very fewer accuracy is found in Gujarati handwritten conjunct characters. Gujarati characters are different than other languages because it does not have Shirolekha over

characters and different types of curves. Gujarati characters are set of 34 consonants, 14 vowels, and 10 numerals [2].

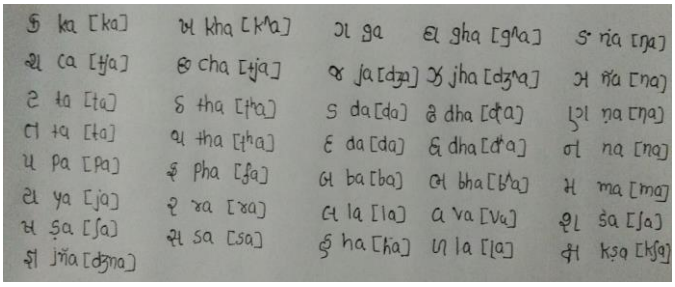

Fig -1: Consonants

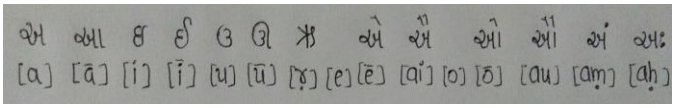

Fig -2: Vowels

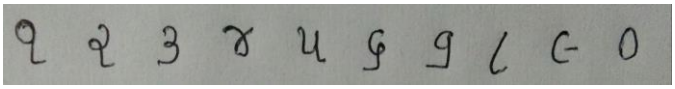

Fig -3: Digits

2.1. Conjunct Characters

Conjunct Characters are the combination of more than one character. Conjunct character recognition is a very problematic research area because writing styles of every person is different. Conjunct Character Recognition is not a difficult task for humans, but for a machine it is very difficult to identify.

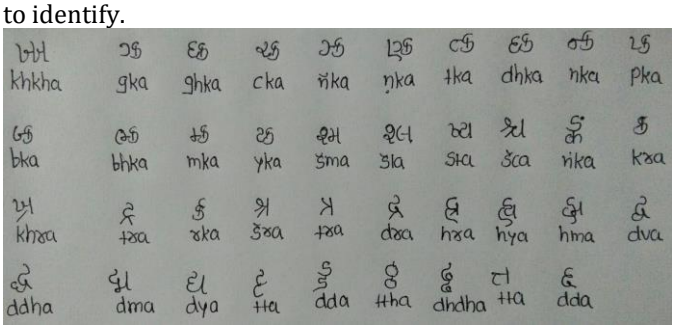

Fig -4: Conjunct Characters

Example: સ્વચ્છ

3. Steps for HCR

Whenever a document is thought for recognition, there are enumerable factors involved herewith. Firstly, the document is scanned so that the text on paper becomes the image on computer. Then this image is pre-processed and then converted into either machine-editable format of just recognized as the set of characters or might be converted into some other script [3].

Pre-processing steps :

i. Image Acquisition

ii. Binarization of scanned image

iii. Removal of Noise from scanned image

iv. Thinning of binarized image

v. Skew detection and correction of scanned image

vi. Segmentation of image

There are many problems encountered in the segmentation procedure .

• Problems with conjunct character segmentation.

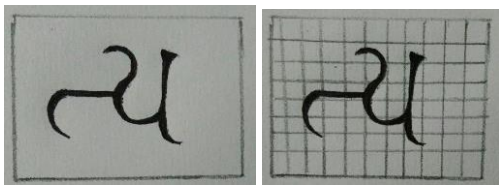

Fig -5: Conjunct Character Segment

• Problems in Line Segment.

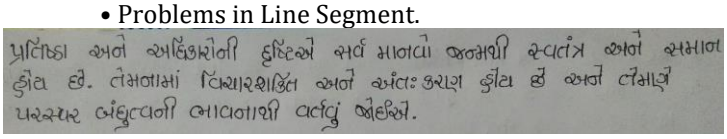

Fig -6: Line Segment

• Problems in Word segmentation.

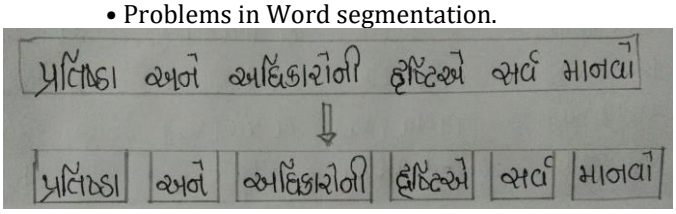

Fig -7: Word Segment

• Problems in Character segmentation.

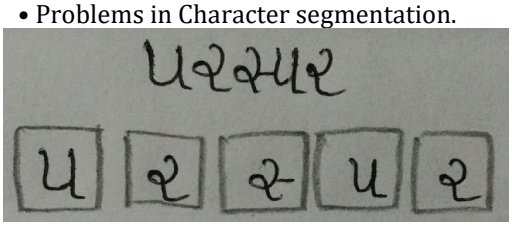

Fig -8: Word Segment

vii. Feature Extraction Techniques

viii. Recognition on the basis of Classifiers

4. CONCLUSIONS

Handwritten character recognition is very difficult task, especially for conjuncted characters of Gujarati script. To get an accurate recognition of conjuncted characters, we must

have to focus on Segmentation, Thinning, Binarization, Feature Extraction.

REFERENCES

[1] Vishal A. Naik, 30 Sept,2018, "Online Handwritten Gujarati Numeral Recognition Using Support Vector Machin". E-ISSN: 2347-2693], Vol.-6, Issue-9, Sept. 2018.

[2] 2) Shailesh Chaudhari and Dr. Ravi Gulati, "Segmentation Problems in Handwritten Gujarati Text". ISSN: 2278-0181, Vol. 3 Issue 1, January -2014.

[3] 3) Ayush Purohit, Shardul Singh Chauhan, "A Literature Survey on Handwritten Character Recognition" ,Vol. 7 (1) , 2016.

[4] 4) Ankur Kumar Aggarwal, Aman Kumar Aggarwal ," Devanagari Script Conjunct Characters Segmentation Based on Characters Structural Properties by Horizontal Projection", Vol.4 Issue 2, April-June 2013.