

A Methodology to Analyze Handwritten MODI Character in Digital Image

Dr. Kirti Nilesh Mahajan¹, Niket P. Tajne²

Professor, Computer Application, Bharti Vidyapeeth Deemed to be University
Institute of Management & Entrepreneurship Development, Pune, India¹

Research Scholar, Computer Application, Bharti Vidyapeeth Deemed to be University
Institute of Management & Entrepreneurship Development, Pune, India²

Abstract: This paper presents the comprehensive analysis of handwritten character Recognition of MODI script. The conversion of a handwritten character into the machine-readable format is nothing but a Handwritten Character Recognition Process. MODI document is available in many fields like Art, Commerce, Social Studies, Health Care, etc. The handwritten character of MODI script is cursive in nature, because of this reason the difficulties level is increase to identify the correct MODI handwritten characters. This research paper showcases the proposed methodology to analyze the handwritten MODI character from a digital image. This proposed methodology deals with the MODI character within the image and it will be used for MODI script transliteration. The latest research in Handwritten Character Recognition has used different types of method, classifiers, and features to reduce the complexity of recognizing handwritten text. In addition, this research paper analyses the different segmentation techniques of Image Processing, which is useful in character identification and mapping. The current paper is categorized into three sections. In First Section is the Introduction of MODI script, Steps of Second Section is the proposed methodology for MODI character Recognition. And in Third Section the Conclusion based on proposed methodology is mention.

Keywords: Handwritten Character Recognition, MODI Script, Digital image, Pattern Recognition Techniques, Character segmentation Techniques.

I. INTRODUCTION

In last few decades, the use of digital handwritten documents has been attracting the many fields like translation/transliteration, analysis of character, data security, pattern recognition, emotional computing and many more. It has grown from the dream to reality. The different application read the handwritten characters from many sources like image, word or pdf documents very accurately. But the most important aspect to learn different handwritten documents is to gain different methodology, strategy, which was used by many researchers those who was worked on the handwritten documents and still the work is going on. In this proposed research paper the basic ideas for analysing the Digital Handwritten documents and methodology has collected which is use to read the particular handwritten character. The need of analysing this methodology is that it may use for transliterating the ancient script into the primary language of the particular region. In this current section of the research paper the introduction of the Ancient script like MODI script has described, also different steps of pattern recognition and types of handwritten character segmentation has discussed

I.A MODI SCRIPT

MODI script is a Cursive type of writing which is type of the Devanagari script. MODI script documents reveal all significance of not only social, political, cultural but also in artistic activity of the past. The use of MODI manuscripts is primarily reflect in Marathi language and Gujarati language which principally used in western part of India states Maharashtra and Guajarati respectively. Also the essence of MODI script is reflect in Urdu, Kannada, and Hindi [2]. MODI script has 32 vowels and vowels Diacritics, 34 consonant letters and 10 Numerals [3][4].

अ	आ	इ	उ	ए	ऐ	ओ	औ
a	ā	i	u	e	ai	o	au
क	का	कि	कु	के	कै	को	कौ
घ	घा	घी	घु	घे	घै	घो	घौ
gha	ghā	ghi	ghu	ghe	ghai	gho	ghau
च	चा	ची	चु	चे	चै	चो	चौ
ca	cā	ci	cu	ce	cai	co	cau

Fig 1. Vowels and Vowels Diacritics

क	ख	ग	घ	ङ	च	छ	ज
ka	kha	ga	gha	ṅa	ca	cha	ja
झ	ञ	ट	ठ	ड	ढ	ण	त
jha	ña	ṭa	ṭha	ḍa	ḍha	ṇa	ta
थ	द	ध	न	प	फ	ब	भ
tha	da	dha	na	pa	pha	ba	bha
म	य	र	ल	व	श	ष	स
ma	ya	ra	la	va	śa	ṣa	sa
ह	ळ						
ha	ḷa						

Fig 2. MODI Consonants

०	१	२	३	४	५	६	७	८	९
0	1	2	3	4	5	6	7	8	9

Fig 3. MODI Numerals

MODI script writing style is cursive. Which has different section, one character of MODI script is written in four lines. These line are called as Rafer Line or Topmost Line (Urdhwarekha), Head Line or Start/End (Shirolekha), Base Line or Foot Line (Padarekha), Rukar Line or Extreme Bottom Line (Talarekha). Following Diagram show the one word written in MODI Script which is lies in between four Lines[1][5][6].



Fig. 4. MODI Script Character Writing Structure

I.B STEPS OF PATTERN RECOGNITION TECHNIQUE

Pattern Recognition is the science of categorization of acquired input in to its identifiable, which have significant features, or attributes of the data. Pattern Recognition is about presumption the unidentifiable attributes of an observation like zero or one, black or white, ON or OFF. This characteristic of Pattern Recognition is helpful in ancient script character recognition. The Pattern Recognition is basically related with the description and analysis of different values taken from any physical or logical processes. Like accumulating raw data from different sources and take predefined action or proposed action.

The Pattern Recognition Technique has sub steps like Data acquisition, Processing (Pre and Post), Data Segmentation, Feature Extraction, classification of data. These all steps are respectively related to input section, processing section and output section. The following diagram shows the all steps of Pattern Recognition Technique [7][8].

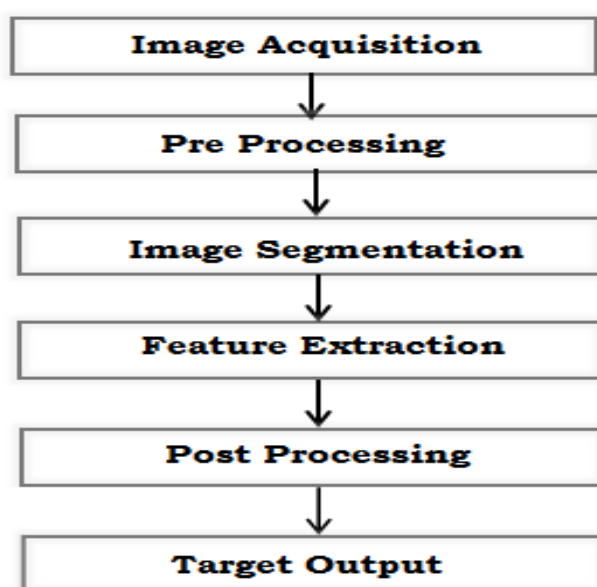


Fig. 5. Steps of Pattern Recognition Technique

I.C SEGMENTATION TECHNIQUES

The Segmentation is most significant section of the Digital image processing and it is widely used in character identification. Also in the Digital Image processing, text extraction is crucial phase to analyse the handwritten text. Generally, the segmentation technique helps to identify the cursive type of handwritten characters by applying various segmentation techniques. The Image segmentation is depends on pixel arrangement of the characters. The final expected output based on handwritten text images can be achieves through the pass different levels of text segmentations. These levels are line segmentation, word segmentation and character segmentation. The sequence of these segmentation levels can be customized according to the demand of the application. The purpose and demand of all these segmentation levels are different. The following diagram show the levels of segmentation with its purpose.

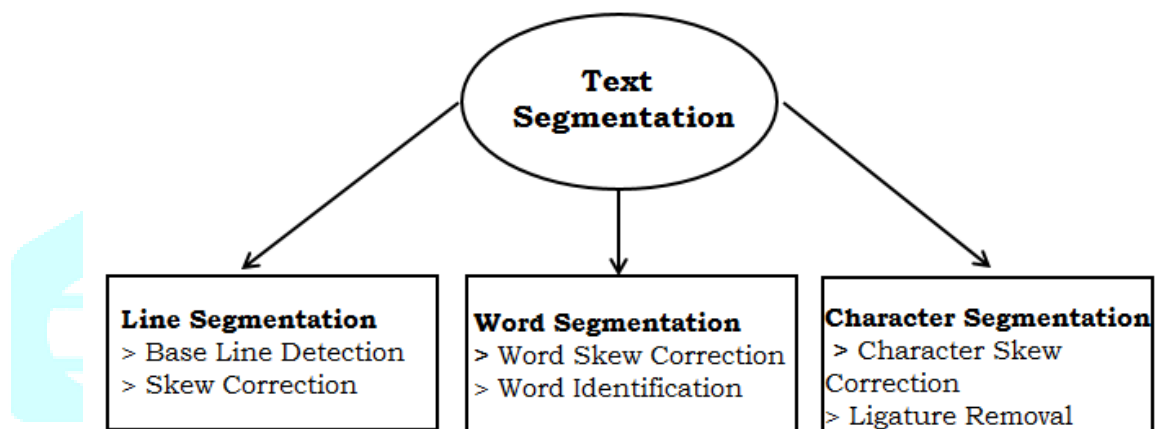


Fig. 6. Levels of Text Segmentation with purpose

I.C.1 LINE SEGMENTATION

Line segmentation is Initial and most significant level of text based image segmentation. Through the line segmentation, not only the unlevelled text character is arranged properly but also it can be useful for the identification of character and translate that character in to subsequent languages or script. In line segmentation, technique the scanning is started from left to right and top to down. Because of horizontal scanning the intensity of each and every pixel is tested. After intensity testing the text character are, dissect in to the small groups. The individual group of the character is represented the individual content of the entire image. These contents of particular text character can be projected vertically or horizontally. The sub methods of the line segmentation are Baseline detection and skew correction.[9,10]

I.C.2 WORD SEGMENTATION

After the completion of line segmentation, the next level is Word Segmentation. Word Segmentation do the separation of the string of text of character into small parts and this is a primary character of that separated character. The whole scanning is starting from the left side and ends on the right side of handwritten documents. Simultaneously it is also scan the pixel row by pixel row with the proper intensity of the pixel. In the word segmentation, two sub-methods are considered in Text-based Word Segmentation steps are word skew correction and Word Identification that are responsible for

correcting the angle of handwritten documents and identify the correct word from the Respective database.

In addition, the word segmentation step of text segmentation deals with two procedure. First procedure is Distance Computation, which is computing the distances of adjacent components in the text line images. Second procedure is the represent the classification of previously computed distances as either inter-word distances or inter-character distances.[11,12]

I.C.3 CHARACTER SEGMENTATION

At the last stage of the Text segmentation, the text has segmented based on the characters hence it is referred as character segmentation. The overall procedure of the Character segmentation is work similarly like word segmentation. This stage reveals the character skew correction which is correct the angle of the handwritten character and Ligature Removal is responsible for the Removing the join character.[10]

II. METHODOLOGY FOR MODI CHARACTER RECOGNITION

The overall methodology of MODI script character recognition is separated into seven different steps. The fig no. 7 illustrates the working of overall methodology of MODI script character recognition diagrammatically. The Whole methodology will start from the physical environment. In Physical environment phase, the actual system is in the ideal state, which is ready to take the input from specific ways. The second phase of a methodology will responsible for acquiring the MODI script handwritten document, which has input files. The format for the input files are any one of type images like .jpg, .jpeg, and .png has been taken.

While inputting the files it may have a chance, the inputted files have less quality of character or line. So the third block of the proposed block diagram of ODI script character recognition is answerable to the eliminating the variability that is inherent in printed words. It eliminating the overlapping characters, overlapping lines, neither parallel nor at right angles to a specified or implied lines or characters, background noise, etc.

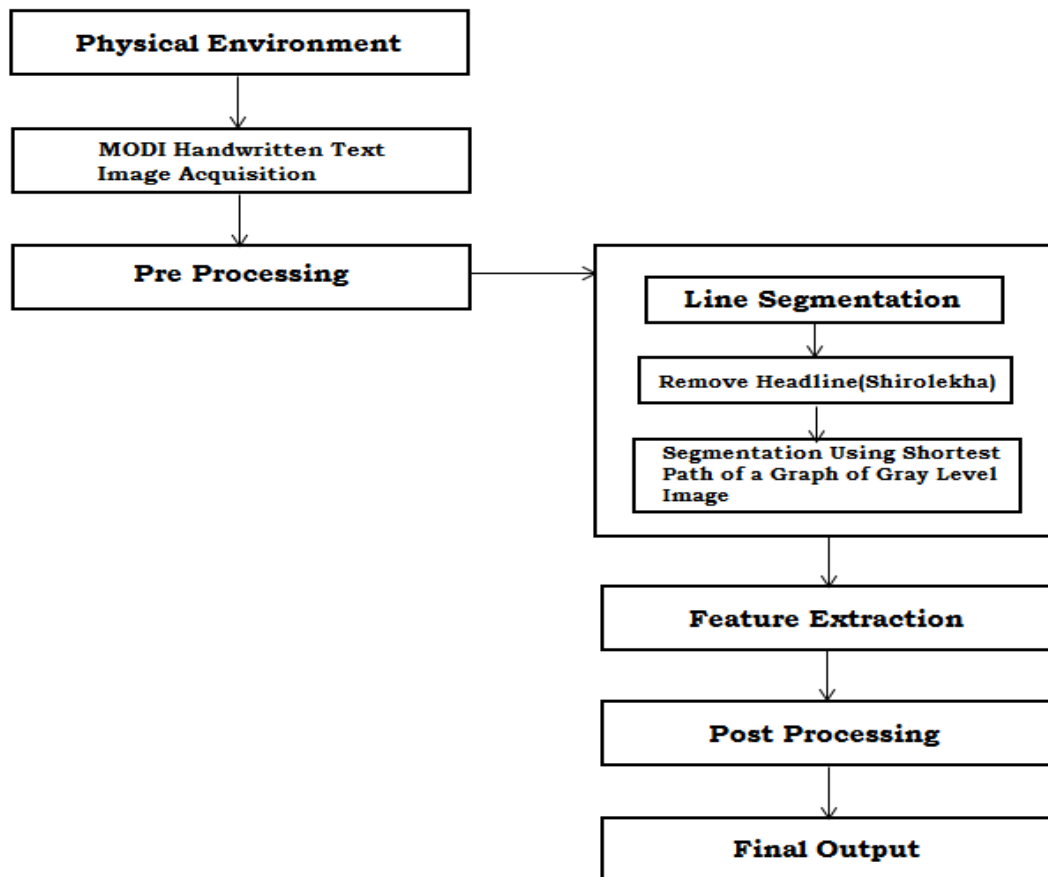


Fig. 7. Basic Block Diagram of MODI Character Recognition

The fourth block of a proposed block diagram of MODI Character Recognition is Segmentation. In the proposed methodology line Segmentation is used for recognizing the MODI character. The writing style of MODI character is cursive and in every word of MODI character the Headline is common this Headline is called as Shirolekha[1]. After segmented the Headline of single MODI word the next segmentation is prepared through the shortest path of a graph of the grey level image is generated.

The Fifth block of a Described Methodology is the Feature Extraction. This block uses the segmented data for the match with the existing data, which is mapped and transfer to the next step. After the Feature Extraction Post Processing block of the proposed methodology is there. This block checks the final transliteration of the MODI script remove the noise from the data if it has. Moreover, in the final block, the output is display on the screen.

III. CONCLUSION

In this research, paper as discussed the proposed methodology for analysing the MODI character from Digital images brings a conclusion that the proposed methodology uses the segmentation technique like line segmentation, which appropriately segmented the MODI characters. After segmentation, it applies the shortest path of a graph of the grey level image and transfers the segmented character for data mapping to the feature extraction section of a methodology. Because of the writing style of writing the MODI alphabets in the proposed research paper the line segmentation method of separation of MODI character is chosen. The current stage of research methodology is portrayed that the analysing of MODI characters from the Digital images will be performed well for the limited characters of MODI script. There is a scope for improving the features in

proposed methodology. Hence at the initial stage, a simple yet effective methodology for the analysing and segmentized the MODI character from the digital images have been delineated.

REFERENCES

- [1] Reviving Modi-script [Online]. Available: http://www.typoday.in/2014/spk_papers14/rajendrathakre-typo14.pdf
- [2] D. N. Besekar, R. J. Ramteke "Study for Theoretical Analysis of Handwritten MODI Script – A Recognition Perspective" International Journal of Computer Applications, 0975 – 8887, Feb-2013.
- [3] Modi Alphabet [Online]. Available: <https://www.omniglot.com/writing/modi.htm>
- [4] Shubhangi Bhatambrekar, Niket Tajne"Context Free Grammar (CFG) for MODI Script", International Conference on Computer Science, Data Mining and Mechanical Engg, April 2015.
- [5] Girish Dalvi, "Anatomy of Devanagari Typefaces", [Online]. Available: <http://www.idc.iitb.ac.in/resources/dt-jan-2009/Anatomy%20of%20Devanagari.pdf>
- [6] Rajendra Bhimraoji Thakre, Typography and Culture Reviving Modi-script [Online]. Available: http://www.typoday.in/2014/spk_papers14/rajendrathakre-typo14.pdf
- [7] Rafael C. Gonzalez, "Digital Image Processing Third Edition", University of Tennessee Richard E. Woods MedData Interactive [Online]. Available:http://web.ipac.caltech.edu/staff/fmasci/home/astro_refs/Digital_Image_Processing_3rdEd_truncated.pdf
- [8] Pattern Classification, [Online]. Available:http://www.byclb.com/TR/Tutorials/neural_networks/ch1_1.htm
- [9] Namrata Dave, "Segmentation Methods for Hand Written Character Recognition",International Journal of Signal Processing, Image Processing and Pattern Recognition Vol. 8, No. 4 (2015), pp. 155-164, 2015.8.4.14.
- [10] Gupta Mehula*, Patel Ankitab, Dave Namratac, Goradia Rahuld, and Saurin Shethe, "Text Based Image Segmentation Methodology", 2nd International Conference on Innovations in Automation and Mechatronics Engineering, ICIAME 2014,
- [11] AMANDEEP KAUR, SEEMA BAGHLA, SUNIL KUMAR, "Study of Various Character Segmentation Techniques for Handwritten Off-Line Cursive Words : A Review", International Journal of Advances in Science Engineering and Technology, ISSN: 2321-9009 Volume- 3, Issue-3, July-2015.
- [12] G. Louloudis1 , B. Gatos , I. Pratikakis , C. Halatsis Line And Word Segmentation of Handwritten Documents [Online]. Available: <http://users.iit.demokritos.gr/~bgat/cr1115.pdf>.