DESIDOC Journal of Library & Information Technology, Vol. 39, No. 2, March 2019, pp. 109-116, DOI : 10.14429/djlit.39.2.13615

Analysis of Segmentation Methods for Brahmi Script

Ajay P. Singh* and Ashwin Kumar Kushwaha

Department of Library and Information Science, Banaras Hindu University, Varanasi – 221005, India
**E-mail: apsingh_73@yahoo.co.in*

ABSTRACT

Segmentation is an important step for developing any optical character recognition (OCR) system, which has to be redesigned for each script having, non-uniform nature/property. It is used to decompose the image into its sub-units, which act as a basis for character recognition. Brahmi is a non-cursive ancient script, in which characters are not attached to each other and have some spacing between them. This study analyses various segmentation methods for different scripts to develop the best suitable segmentation method for Brahmi. MATLAB software was used for segmentation purpose in the experiment. The sample data belongs to Brahmi script-based 'Rumandei inscription'. In this paper, we discuss a segmentation methodology for distinct components, namely text lines, words and characters of Rumandei inscription, written in Brahmi script. For segmenting distinct components of inscription different approach were used like horizontal projection profile, vertical projection profile and Relative minima approach. This is fundamental research on an inscription based on Brahmi script, which acts as a foundation for developing a segmentation module of an OCR solution/system of similar scripts in future. Information search and retrieval is an important activity of a library. So, to ensure this support for digitised documents written in ancient script, their character recognition is mandatory through the OCR system.

Keywords: Digitisation; Projection profile; Segmentation process; Connected character; Line segmentation; Word segmentation; Character segmentation; Brahmi script

1. INTRODUCTION

Brahmi and its successor scripts were predominant in Indian history, literature written in these scripts integrates historical and cultural knowledge of Indian peninsula. The recognition of various characters of Brahmi has not received proper attention as Latin, Devnagri, Arabic, Kannada, Oria, or Bangla scripts have got[1-7]. In the domain of pattern recognition, the most difficult and demanding task is character recognition. A state of excellence is achieved for recognising printed characters but for handwritten characters, there still exists a gap[8]. Research on character recognition for Indic scripts is on progression and at present, there is no solution that solves the problem correctly and efficiently[9]. Every optical character recognition (OCR) system involves a specific type of segmentation module, where decisions are taken for dividing a document image text into its fundamental parts like text line, words, characters, or sub-characters. These fundamental parts further act as basic units for the character recognition process. The accuracy of line/word/character segmentation affects the results of character recognition accordingly[10]. Therefore, it is very important to segment the document image in such a manner, where syntactical and grammatical issues may accordingly address while character recognition. The segmentation task encounters a number of challenges while practical implementation. In text line segmentation approach, major challenges include the availability of noise-free images of document, difference in the skew angle among various text lines and also between the text line and page, overlapping among the words and another text line, overlapping of characters within the word, and finally non-uniform spacing of words in the text.

Here we have focused on the identification of the best suitable segmentation approach for segmenting lines, words, and characters of 'Rumandei inscription' of Brahmi script. The primary source of Rumandei inscription for the present research is 'Rumandei minor pillar edict' of Mauryan king Asoka, found in Lumbini region of Nepal. Besides this primary source, we have also gathered secondary images of Rumandei inscription from Brahmi script based corpus of Alexander Cunningham. Images from multiple sources have been used for ensuring generalisation of the segmentation process of identified inscription[11].

We have applied the Projection profile based segmentation approach. In the text line segmentation approach, a projection profile of an image is taken for analysing separate lines, words, and characters. Projection profile of an image refers to the running sum of the pixels in a particular direction[12]. In this approach, for line segmentation, a horizontal projection profile of text image is taken, and corresponding to it a graph is plotted, which was analysed for the segmentation purpose. For, word and character segmentation, a verticle projection profile of a segmented line and word is taken respectively,

Received : 17 September 2018, Revised : 04 February 2019
Accepted : 05 February 2019, Online published : 11 March 2019

and corresponding to both of it a graph is plotted to perform segmentation of line and word.

2. RELATED WORK

As per our observation, no work has been reported yet on the segmentation of characters incorporating the ancient Brahmi script forms. Although significant work has been reported in the literature on OCR of both Indic and Non-Indic scripts. Various segmentation methods which were investigated are as follows:

2.1 Projection Profile Method for the Line, Words and Character Segmentation

Pal & Sarkar[13] exploited valleys of projection profile for segmenting text document image into lines, which was analysed by counting the number of black pixels in each row. Combination of Vertical projection profile and component labelling methods have been exploited for character segmentation. In Urdu-Nastaleeq, font segmentation of words into ligatures can be done by connected component and baseline using a horizontal profile[14]. Priyanka[10], et al. suggested a technique for segmenting text lines depends on the modified histogram achieved by run-length based smearing. The segmented lines of a text document image were obtained by identifying the valleys in the histogram, which is computed by summing the row-wise pixel values. The peak between two consecutive valleys represents the text line[15]. Urdu words have been segmented into ligatures while using two methods- connected component labelling and centroid to centroid[16], it reflects multi-tier holistic approach in the process of segmentation.

2.2 Neural Network Based Segmentation Techniques

Since 1990, artificial neural networks are used as a different approach for image segmentation. Characteristics like graceful degradation in the presence of noise, ability to be used in real-time applications and the ease of implementing them with VLSI processors, led to a booming of ANN-based methods for segmentation[17]. Hamad[18] proposed a segmentation technique for Arabic handwritten script that has been used in two phases: Arabic Heuristic Segmenter and Neural base segmentation technique. Xiao & Leedham[19] used a typical approach for segmenting characters, firstly characters are segmented into sub-component on the basis of certain criteria and then over segmented sub-component are merged on the basis of their characteristics. Neural networks are the best way to tackle the problem of over-segmentation. The family of neural networks used for pattern recognition, which is simple and effective to implement is Feed-Forward Back Propagation network[18].

2.3 Segmentation Free Approach

Spotting and segmentation are done simultaneously in segmentation free approach with the help of a sliding window. Feasibility and success of the algorithm are tested by spotting method but the final result is free from the actual spotting method used[20]. For the purpose of applying segmentation free approach to *Devnagri* scripts, long short-term memory (LSTM) networks are suitable due to their good context-aware processing. LSTM networks are kind of Recurrent Neural Networks and wisely used in unsegmented character recognition[21]. Premaratne & Bigun[22] proposed a novel approach for printed character recognition using linear symmetry, where features are directly used to recognise characters with standard alphabets without performing actual segmentation. Vertical modifiers in the text are identified by edge detection algorithm which uses linear symmetry and skews angle. For Arabic script, various methods based on segmentation free approach were applied for characters and the satisfactory result for character recognition was obtained[3,23-24].

2.4 Hough Transform Approach for Line Segmentation

Hough Transform principle can be utilised in various research areas like skew detection, slant detection, text line segmentation, etc.[25]. Hough Transform is used for the directional segmentation of lines and words from various kinds of images. The Hough image is generated from the binarised edge map of the image and the parameters of the Hough transform are tuned depending on the variety of the images used[26].

2.5 Other Approaches for Character Segmentation

Chaudhuri[27], *et al.* divided character segmentation into three categories as explicit segmentation, implicit segmentation and mixed strategies. Two fuzzy features were used to identify the *Matra* region and potential segmentation point in a script independent character segmentation from word images[28]. Bishnu & Chaudhuri[29] discussed a technique called recursive contour, which explores the range of zonal height, inside which the prominent portion of character exists. Acceptable outputs are obtained if the adjacent characters are not touching. Naz[3], et al. segmented the text image into lines and the contours of ligatures, which were recognised using shape descriptors.

3. ANCIENT INDIAN SCRIPT- BRAHMI

Indus script is the earliest script of ancient India that was written by Indus people on seals, sealing, pottery etc. As Indus was a pictographic script thus, we consider Brahmi as the oldest script of India[30]. Brahmi is derived from the Phoenician text, so the antiquity of this script must be 4th to 5th century B.C.[31]. Firstly, Brahmi script was successfully deciphered by James Prinsep in the 1835. Brahmi script emerged in a finished form in the Mauryan period. Brahmi script does not seem to have resulted from the historical process of evolution. It is an instance of conscious creation of a script at some fixed period[32].

3.1 Characteristics

The Ashokan Brahmi script consists of seven vowels including nasal words and thirty-three consonants including semi-vowels. There is no phase consonant as such, later on, which was indicated by the form of the required consonants letter only[33].

Brahmi inscriptions are found engraved on rocks, pillars, caves and on the slab. Lines found in them seen parallel to each other from left to right. Some sort of punctuation is also observed

by leaving some space between two words[33]. The evidence of writing is found in the form of inscriptions indisputably from the Ashokan period, and goes back to 3rd century B.C. Ashoka engraved his messages in two main scripts of the country- the *Brahmi* and the *Kharosthi*[34].

3.2 Style of Writing

In various Brahmi inscriptions following attributes have been identified regarding their writing style:

- The writings follow ink style, as it is clear from the use of dots in some letters (i.e. Kha-, Ja- and Da-)
- It is seldom that the scripts attempted to achieve quality in most of the cases. They were governed by the original form of letters rather by their own inclination. (i.e. U, E, O, Ka, Ga, Jha, Na, Da, Na, Ba, and Bha)
- There is no abruptness in writing. (Mysore inscription is the only exception)
- Each letter is distinctly formed
- No cursive writing in the way of continuously drawn outlines.

3.3 Engraving Style of the Script

Early Brahmi inscriptions are engraved on rocks, pillars and caves, while one is the carved on a stone slab[35]. While assessing inscriptions like *Rumandei, Sarnath* etc. rock edict; it is observed from Fig. 1 that engraving style differs from hand to hand. Absorbingly, the variance may occur even though the material on which the record is engraved[35]. In the rock edict of Girnar only, we could identify nine forms of 'a' etc. It has 35 characters which are used in a few different styles. Among them, 6 vowels and *anuswara*, 23 consonant, 4 semi-vowels and 2 aspirate have been included[36].

Vowels		
A-	I-	U-
Consonants		
K-	Ga- (Guttural)	Gha-
Ch-	Ja- (Palatal)	Jha-
Ta-	Da- (Lingual)	Dha-
Ta-	Da- (Dental)	Na-
Pa-	Ba- (Labials)	Ma-
Ya-	La- (Semi-vowels)	Va-
	Sa- (Sibilants)	
	Ha- (Aspirate)	

Figure 1. Basic or primary forms of characters of Brahmi.

4. SEGMENTATION PROCESS

Segmentation is a method which determines the elements of an image. When practised with text, segmentation is the process of separating characters or words[37]. In computer vision, breaking of a digital image in some collective segments is known as picture segmentation. More precisely picture segmentation is the process of labelling of every image pixel in such a way that pixels assigned with similar tags have some distinct characteristics with others[38]. Eikvil[37] mentions that most of the OCR methods recognise character individually by segmenting the words into separate characters. Generally, the connected component method is used for the segmentation process, this method is easy to implement unless the characters are non-fragmented and non-touching.

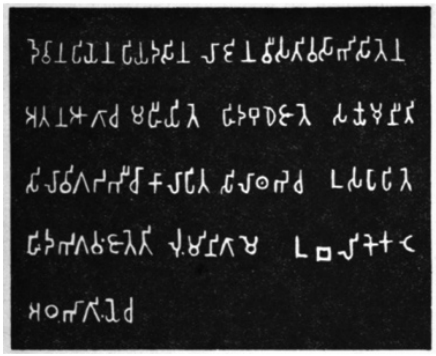

Figure 2: Scanned image.

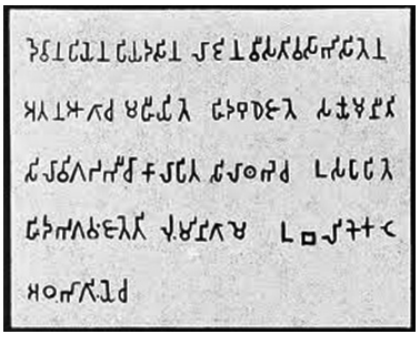

Figure 3. Pre-processed.

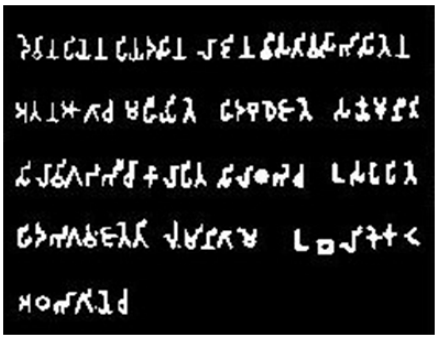

Figure 4. Noise-free image with noise image with reduced noise obtained after pre-processing and binarisation

4.1 Preprocessing

The image of *Rumandei* inscription obtained after scanning may consist of noise (Fig. 2). The obtained image depends on the

resolution power of the camera and thresholding method used. So, this dependency of the image may lead to low recognition rate, which can be wiped out by the preprocessor. Preprocessor helps to smooth the image by applying various process like filling, thinning and normalization (Figs. 3 and 4). The filling is done to remove gaps, breaks and holes while thinning minimises the line width of digitised characters. To achieve characters of uniform slant, size and rotation normalization is done [37].

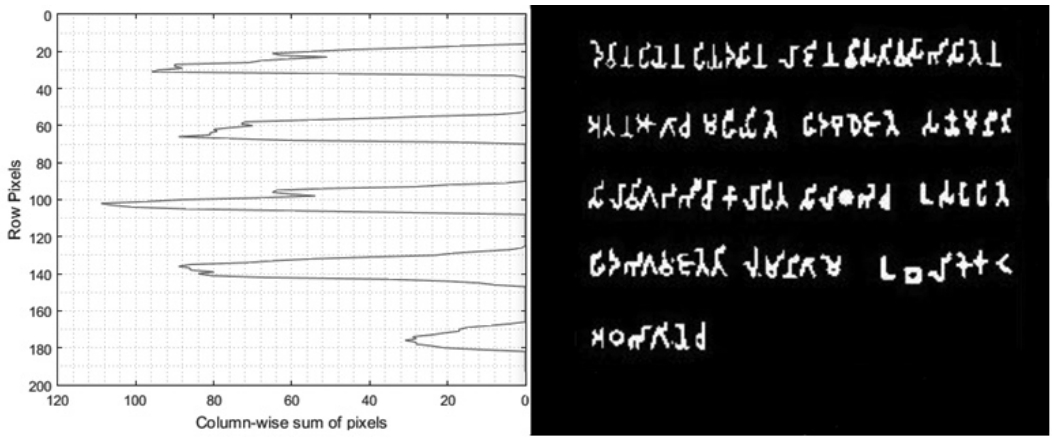

Figure 5. Horizontal projection profile for image segmentation.

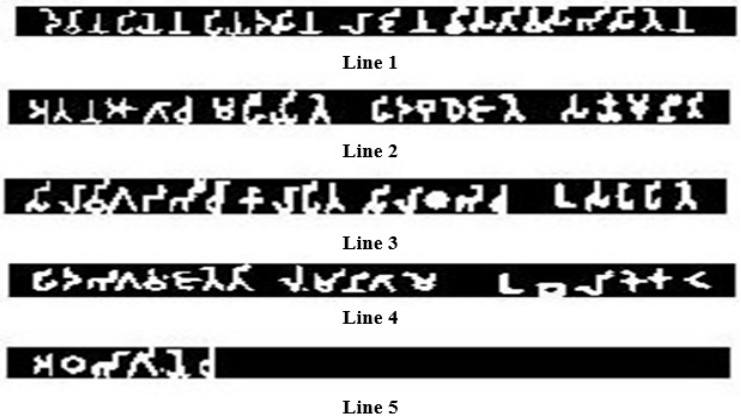

Figure 6. Segmented lines from the image.

5. LINE SEGMENTATION

Line segmentation is done by horizontal projection profile methods[25]. The whole methodology and experiments were performed on MATLAB. Initially, a pre-processed binarised image was taken. Then horizontal projection profile of that binarised image was developed by summing column pixel values. The sum of column pixel values was stored in an array in a matrix form. Now, the matrix of the projection profile was analysed and the pixel of interest was found. Point/Pixel of interest is a starting point from where the projection profile initiates or it is greater than zero till again the projection profile reaches to the value zero. This can be achieved by applying the loop with some logical statement in it. To understand more clearly, a pictorial representation of the above logic is shown in Fig. 5. The valley between two consecutive peaks/crest in Fig. 5, represents the space between two text lines. The peak/crest between two consecutive troughs in Fig. 5, represents the text line. So, points representing lines were cropped out using MATLAB functions and hence the segmented lines are shown in Fig. 6.

6. SEGMENTATION OF LINE INTO WORDS

Word segmentation is done by vertical projection profile methods[25]. Nath, Jelil & Rahul[39] states that Word segmentation method assumes that the spacing among words is always greater than the spacing among the characters. So, if the vertical projection is performed to calculate the corresponding vertical projection profile (Fig. 7), based on the threshold distance (D) among two consecutive zero (Fig. 8). The rupture points of the histogram have been used to recognise the word (Fig. 9).

Here, Line 1 was taken as an input and binarised for further process. Vertical projection profile of the segmented image of Line 1 is taken by summing row pixel values of the binarised image. The sum of row pixel values was stored in an array in a matrix form. Now, the matrix of the projection profile was analysed and the pixel of interest was found. Point/pixel of interest is a starting point from where the projection profile initiates or it is greater than zero for five or more consecutive pixels till again the projection profile reaches to the value zero for five or more consecutive pixels. To understand more clearly, a pictorial representation of the above logic is as shown in Fig. 7. The valley between two consecutive peaks/crest in Fig. 7, represents the space between two characters or words. The peak/crest between two consecutive troughs in Fig. 7, represents the characters or words. So, points representing words are those who have a threshold distance (D) greater than five (D > 5) (Fig. 8). So, these words were cropped out using MATLAB functions and hence the segmented words are reflected in Fig. 9.

7. SEGMENTATION OF WORD INTO CHARACTERS

Character segmentation is done through vertical projection profile methods[13]. Here, Word 1 was taken as an input and binarised for further process. Vertical projection profile of the segmented image of Word 1 is taken by summing row pixel values of the binarised image. The sum of row pixel values was stored in an array in a matrix form. Now, the matrix of the projection profile was analysed and the pixel of interest was found. Point/pixel of interest is a starting point from where the projection profile initiates or it is greater than zero till again the projection profile reaches to the value zero. To understand more clearly, a pictorial representation of the above logic is shown in Fig. 10. The valley between two consecutive peaks/crest in

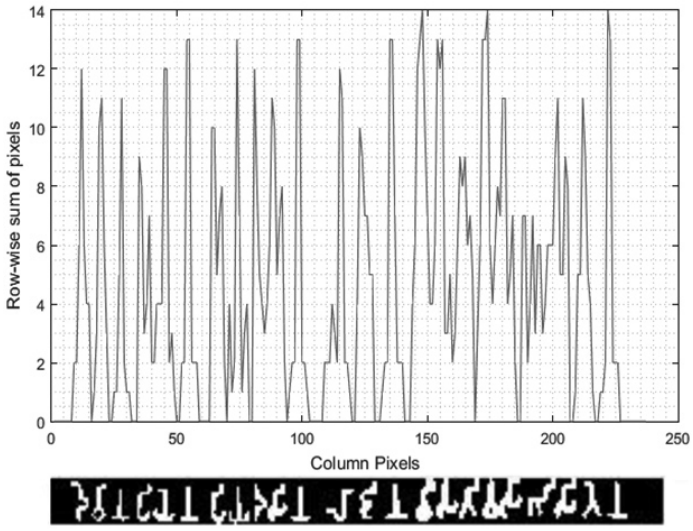

Figure 7. Vertical projection profile for line segmentation.

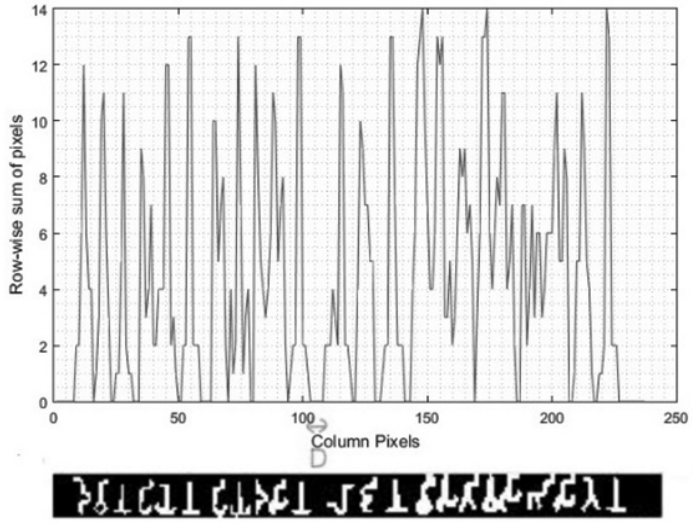

Figure 8. Threshold distance (D) between words in Vertical projection profile of the image.

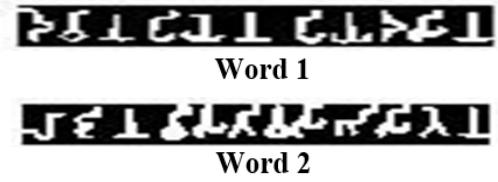

Figure 9. Segmented words from a line.

Fig. 10, represents the space between two characters. The peak/crest between two consecutive troughs in Fig. 10 represents the characters. So, the points representing characters were cropped out using MATLAB functions and hence the segmented word/character is reflected in Fig. 11.

7.1 Segmentation of Connected Characters into Individual Characters

Segmentation of connected characters is done through vertical projection profile methods[13]. In connected character

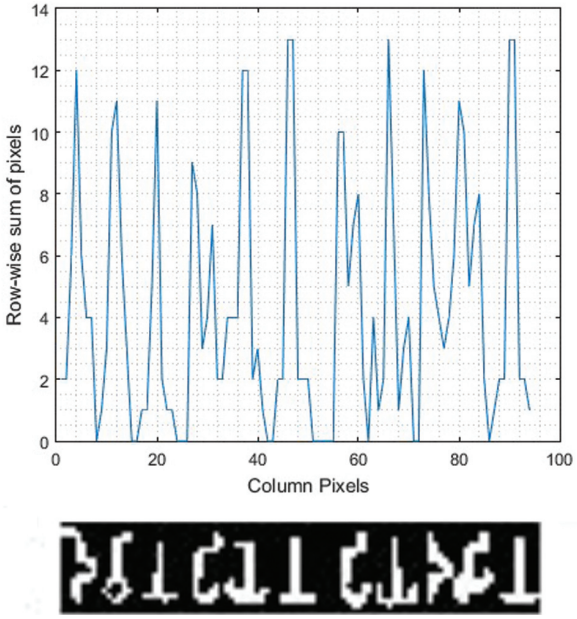

Figure 10. Vertical projection profile for character segmentation.

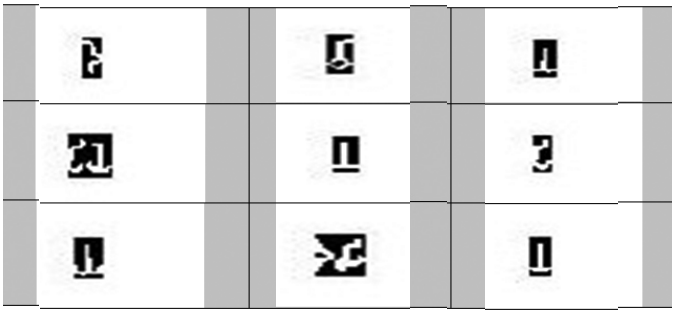

Figure 11. Segmented characters from a word.

segmentation, first of all, there is a need to identify connected characters. It can be achieved by evaluating segmented characters, characters which have a width greater than 9 pixels are the required connected character and hence used as an input. Vertical projection profile of segmented input image is taken by summing row pixel values of the binarised input image. The sum of row pixel values was stored in an array in a matrix form. Now, the matrix of the projection profile was analysed and the pixel of interest was found. Point of interest is a starting point, whenever the first minima of projection profile were found for the overlapped/connected character. So, from this point segmentation of character is done. To understand more clearly a pictorial representation of the above logic is as shown in Fig. 12. The red line denotes the boundary for segmentation for the connected components. So, points representing characters were cropped out using MATLAB functions and hence character segmentation of connected characters occurs. The segmented characters from overlapped character were as shown in tabular form in Fig. 13.

8. DISCUSSION AND CONCLUSION

The data sample belongs to 'Rumandei inscription' of Brahmi script. Figures 6, 9, 11, and 13 shows the output of the proposed method. From the experimental results, it is clear that

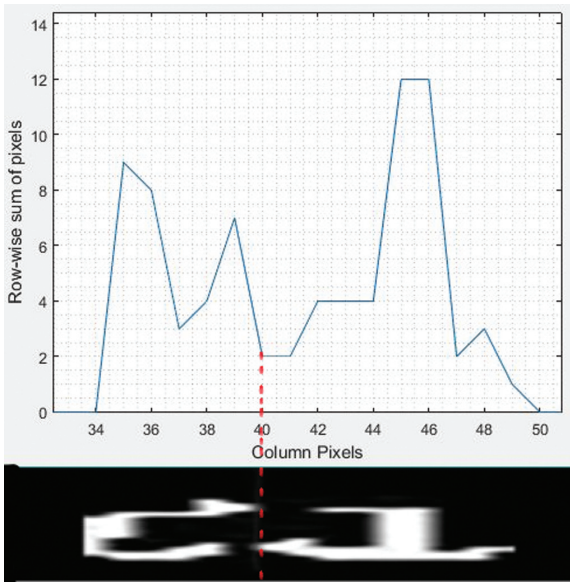

Figure 12. Vertical projection profile for overlapped characters segmentation.

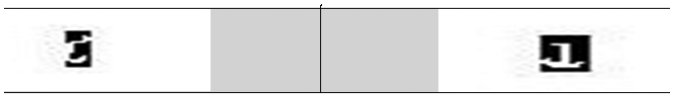

Figure 13. Segmented characters from an over-lapped character.

the method which is proposed is suitable for segmenting the image, lines, words and overlapped characters satisfactorily.

OCR's character segmentation and classifier stage are required to be redesigned for each new script, while other stages can be used as such with certain modification for large classes of languages[1]. So, in this paper, we have discussed various approaches of segmentation with respect to different scripts/languages and then tried to discover the best suitable technique for segmentation of Brahmi script-based 'Rumandei inscription'. However, these results are achieved on 'Rumandei inscription' but this method may not produce a similar satisfactory result for other inscriptions of Brahmi. It is due to the following reasons:

- The difference in the skew angles among various units of an inscription
- Availability of noise-free image, due to various reasons.
- Breaks in the characters, which leads to over-segmentation
- Overlapping of lines, words and characters, which leads to mis-segmentation.

This paper concludes that, at present, we do not have any uniform method for Brahmi script, based text segmentation, and hence, for character recognition of Brahmi text. So, it is an open research problem.

8.1 MATLAB function used

J = imcrop(I) displays the image I in a figure window and creates an interactive Crop Image tool associated with the image. I can be a grayscale image, a true-color image, or a logical array*.

S = sum(A) returns the sum of the elements of A along the first array dimension whose size does not equal 1*.

* Source for the above code is https://in.mathworks.com/help/index.html

REFERENCES

1. Agrawal, M.; Ma, H. & Doermann, D. Generalisation of Hindi OCR using adaptive segmentation and font files. *In* Guide to OCR for Indic Scripts, advances in pattern recognition, edited by V. Govindaraju & S. Setlur. Springer, London, 2009, 181-207. doi: 10.1007/978-1-84800-330-9_10.
2. Preethi, P. & Mamatha, H.R. A review on automation of ancient epigraphical images. *Int. J. Database Theory Appl.,* 2016, **9**(4), 143-150.
3. Naz, S.; Hayat, K.; Razzak, M.I.; Anwar, M.W. & Akbar, H. Arabic script based character segmentation: A review. *In* 2013 World Congress on Computer and Information Technology (WCCIT), 22-24 June 2013, Sousse, Tunisia. 2013. pp. 1-6. doi: 10.1109/WCCIT.2013.6618741.
4. Roy, P.P.; Bhunia, A.K.; Das, A.; Dey, P. & Pal, U. HMM-based Indic handwritten word recognition using zone segmentation. *Pattern Recognition,* 2016, **60**, 1057-1075. doi: 10.1016/j.patcog.2016.04.012.
5. Bag, S.; Harit, G. & Bhowmick, P. Recognition of Bangla compound characters using structural decomposition. *Pattern Recognition*, 2014, **47**(3), 1187-1201. doi: 10.1016/j.patcog.2013.08.026 .
6. Chaudhuri, B.B.; Pal, U. & Mitra, M. Automatic recognition of printed oriya script. *In* Proceedings of sixth International Conference on Document Analysis and Recognition (ICDAR), 13-13 Sept. 2001, Seattle, WA, USA. 2001. pp. 795-799. doi: 10.1109/ICDAR.2001.953897.
7. Pal, U. & Chaudhuri, B.B. Indian script character recognition: a survey. *Pattern Recognition*, 2004, **37**(9), 1887-1899. doi: 10.1016/j.patcog.2004.02.003.
8. Dash, K.S.; Puhan, N.B. & Panda, G. Odia character recognition: A directional review. *Artif. Intell. Rev*., 2017, **48**(4), 473-497. doi: 10.1007/s10462-016-9507-5
9. Kumar, M.; Jindal, M.K.; Sharma, R.K. & Jindal, S.R. Character and numeral recognition for non-Indic and Indic scripts: a survey. *Artif. Intell. Rev*., 2018, **40**, 1-27. doi: 10.1007/s10462-017-9607-x.
10. Priyanka, N.; Pal, S. & Mandal, R. Line and word segmentation approach for printed documents. *IJCA, Special Issue on RTIPPR*, 2010, **4**, 30-36. https://www.ijcaonline.org/specialissues/rtippr/number1/973-96. (Accessed on 1 July 2018).
11. Cunningham, A. Corpus inscriptionum indicarum: Inscriptions of Ashoka. Office of the superintendent of

government printing, Calcutta, 1877.
12. Jha, S. What is horizontal and vertical projection of image in MATLAB? 2017. https://www.quora.com/What-is-horizontal-and-vertical-projection-of-image-in-MATLAB (accessed on 29 July 2018).
13. Pal, U. & Sarkar, A. Recognition of printed Urdu script. *In* Seventh International Conference on Document Analysis and Recognition (ICDAR 2003), 6-6 Aug. 2003, Edinburgh, UK. 2003. pp. 1183-1187. doi: 10.1109/ICDAR.2003.1227844.
14. Shah, Z.A. Ligature based optical character recognition of Urdu-Nastaleeq font. *In* International Multi Topic Conference, 2002, 27-28 Dec. 2002, Karachi. 2002. pp. 25-25. doi: 10.1109/INMIC.2002.1310132.
15. Pal, U. & Chaudhuri, B.B. Automatic separation of words in multi-lingual multi-script Indian documents. *In* Proceedings of the Fourth International Conference on Document Analysis and Recognition. Vol 2, 18-20 Aug. 1997, Ulm, Germany. 1997. pp. 576-579. doi: 10.1109/ICDAR.1997.620567.
16. Husain, S.A. A multi-tier holistic approach for Urdu Nastaliq recognition. *In* International Multi Topic Conference, 2002. INMIC 2002, 27-28 Dec. 2002, Karachi. 2002. pp. 84-84. doi: 10.1109/INMIC.2002.1310191.
17. Amza, C. A review on neural network-based image segmentation techniques, 2015. https://www.researchgate.net/publication/228873725_A_REVIEW_ON_NEURAL_NETWORK-BASED_IMAGE_SEGMENTATION_TECHNIQUES (accessed on 30 July 2018).
18. Al Hamad, H.A. Neural-based segmentation technique for Arabic handwriting scripts. *In* 21st International Conference on Computer Graphics, Visualization and Computer Vision, 2013, Czech Republic. 2013. pp. 9-14. https://www.researchgate.net/publication/311440157 (accessed on 20 July 2018).
19. Xiao, X. & Leedham, G. Knowledge-based English cursive script segmentation. *Pattern Recognition Lett*., 2000, **21**(10), 945-954. doi: 10.1016/S0167-8655(00)00049-0.
20. Ball, G.R.; Srihari, S.N. & Srinivasan, H. Segmentation based and segmentation-free methods for spotting handwritten Arabic words. *In* 10th International Workshop on Frontiers in Handwriting Recognition, Oct. 2006, La Baule, France. 2006. https://hal.inria.fr/inria-00112708 (accessed on 31 July 2018).
21. Karayil, T.; Ul-Hasan, A. & Breuel, T.M. A segmentation-free approach for printed Devanagari script recognition. *In* 13th International Conference on Document Analysis and Recognition (ICDAR), 23-26 Aug. 2015, Tunis, Tunisia. 2015. pp. 946-950. doi: 10.1109/ICDAR.2015.7333901.
22. Premaratne, H.L. & Bigun, J. A segmentation-free approach to recognise printed Sinhala script using linear symmetry. *Pattern Recognition*, 2004, **37**(10), 2081-2089. doi: 10.1016/j.patcog.2004.01.021.
23. Javed, S.T.; Hussain, S.; Maqbool, A.; Asloob. S.; Jamil, S. & Moin, H. Segmentation Free Nastalique Urdu OCR. *Int. J. Comput. Inf. Eng*., 2010, **4**(10), 1514-1519. doi: 10.5281/ZENODO.1080342.
24. Sattar, S.A.; Haque, S.; Pathan, M.K. & Gee, Q. Implementation challenges for Nastaliq character recognition. *In* Wireless Networks, Information Processing and Systems. IMTIC 2008. Communications in Computer and Information Science, vol 20, edited by D.M.A. Hussain; A.Q.K. Rajput; B.S. Chowdhry & Q. Gee, Heidelberg, Berlin, 2008, 279-285. doi: 10.1007/978-3-540-89853-5_30.
25. Dongre, V.J. & Mankar, V.H. Devnagari document segmentation using histogram approach. *Int. J. Comput. Sci. Eng. Inf. Technol*., 2011, **1**(3), 46-53. doi: 10.5121/ijcseit.2011.1305.
26. Saha, S.; Basu, S.; Nasipuri, M. & Basu, D. A Hough transform based technique for text segmentation. *Journal Computing*, 2010, **2**(2), 134-141. https://arxiv.org/ftp/arxiv/papers/1002/1002.4048.pdf. (accessed on 20 June 2018).
27. Chaudhuri, A.; Mandaviya, K.; Badelia, P. & Ghosh, S.K. Optical character recognition systems. *In* optical character recognition systems for different languages with soft computing, edited by A. Chaudhuri; K. Mandaviya; P. Badelia & S.K. Ghosh. Springer International Publishing, Cham, Switzerland, 2017, 9-42. doi: 10.1007/978-3-319-50252-6_2.
28. Sarkar, R.; Malakar, S.; Das, N.; Basu, S. & Nasipuri, M. A script independent technique for extraction of characters from handwritten word images. *Int. J. Comput. Appl.*, 2010, **1**(23), 83-88. (accessed on 17 January 2019).
29. Bishnu, A. & Chaudhuri, B.B. Segmentation of Bangla handwritten text into characters by recursive contour following. *In* Proceedings of the Fifth International Conference on Document Analysis and Recognition. ICDAR 99, 22-22 Sept. 1999, Bangalore, India. pp. 402-405. doi: 10.1109/ICDAR.1999.791809.
30. Marshall, J. Mohenjo-Daro and the Indus Civilization. Asian Educational Services, New Delhi, 1996.
31. Burnell, A.C. Elements of South Indian Palaeography. Asian Educational Services, New Delhi, 1994.
32. Buhler, G. Indian Paleography. Munshiram Manoharlal, Delhi, 2004.
33. Pandey, R. Indian Paleography. Motilal Banarasi Das, New Delhi, 1952.
34. Dani, A.H. Indian Paleography. Clarendon Press, Oxford, 2015.
35. Upasak, C.S. History and Paleography of Mauryan Brahmi Script. Nav Nalanda Vihar, Patna, 1960.
36. Verma, T.P. Brahmi Lipi Ka Udhava Aur Vikash. Ramanand Vidhya Bhawan, New Delhi, 1998.
37. Eikvil, L. OCR - Optical Character Recognition, 1993. https://www.nr.no/~eikvil/OCR.pdf (accessed on 16 Sept. 2019).
38. Janani, G.; Vishalini, V. & Kumar, P.M. Recognition and

analysis of Tamil inscriptions and mapping using image processing techniques. *In* 2016 Second International Conference on Science Technology Engineering and Management (ICONSTEM), 30-31 March 2016, Chennai, India. 2016. 1-8.
doi: 10.1109/ICONSTEM.2016.7560947.

39. Nath, K.; Jelil, S. & Rahul, L. Line, word and character segmentation of Manipuri machine printed text. *In* 2014 Sixth International Conference on Computational Intelligence and Communication Networks, 14-16 Nov. 2014, Bhopal, India. 2014. 203-206.
doi: 10.1109/CICN.2014.55.
40. Naz, S.; Umar, A.I.; Shirazi, S.H.; Ahmed, S.B.; Razzak, M.I. & Siddiqi, I. Segmentation techniques for recognition of Arabic-like scripts: A comprehensive survey. *Educ. Inf. Technology*, 2015, **21**(5), 1225-1241.
doi: 10.1007/s10639-015-9377-5.

CONTRIBUTORS

Dr Ajay P. Singh holds a PhD in Library and Information Science. Presently, he is working as Professor in the Department of Library and Information Science, Banaras Hindu University, Varanasi. He had research experience in OCR, digital preservation, cloud computing applications etc. He has published around 100 research paper in various national and international sources. These results show the follow up research activities of the previously completed UGC Major Project "Designing and Development of an Expert System for Character Recognition of Indian Manuscripts written in Brahmi Script".

Mr Ashwin, holds Master degree in Library and Information Science from Banaras Hindu University, Varanasi. He is availing UGC Junior Research Fellowship and presently doing PhD degree from Banaras Hindu University. He is doing research on 'OCR for ancient Indian scripts'. This paper shows some of the research results of it. He has completed a number of short-term courses on the present area of research.