ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 5, Issue 1, pp: (33-39), Month: January - March 2017, Available at: www.researchpublish.com

Handwritten Devanagari Compound Character Recognition: A Survey

Pooja K. Araj[1], Prof. P. E. Ajmire[2]

[1]M.Sc. II Sem IV Comp. Sci., [2]HOD (Comp. Sci. & Application).
[1]G. S. Sci., Arts & Commerce College, Khamgaon Dist. Buldana (M.S).444303, INDIA
[2]Head, Department of Computer Science & Application, G. S. Sci., Arts & Commerce College, Khamgaon, Dist. Buldana (M.S).444303, INDIA

Abstract: **Compound character recognition of Handwritten Devanagari is one of the challenging tasks due to its complexity as compare to many other scripts. Compound characters itself complex in structure. It is written with combination of two or more characters. The character may be formed with different sequence of combinations of basic characters, such as vowels and consonants. The recognition of compound characters makes this task more challenging to the researchers. The frequency of occurrence of compound characters in Marathi language is more as compared to other languages derived from devanagari script. The various researchers used different classification techniques such as Neural Network, Soft Computing, Seventh Central Moment, Multiclass SVM Classifier with orthogonal moment, wavelet transformation etc. This paper deals with the study of different classification techniques used for compound character recognition of handwritten Devanagari compound character.**

Keywords: **Devanagari, Compound Character.**

I. INTRODUCTION

India is a multi-lingual and multi-script country comprising of eleven different scripts. Devanagari is third most widely used script, used for several major languages such as Hindi, Sanskrit, Marathi and Nepali and is used by more than 500 million people [1]. Handwritten character recognition for Indian script is quite challenging task for the researchers. This is due the various characteristics of these scripts like their large character set, complex shape, presence of modifiers, presence of compound characters and similarity between characters [2, 3, 4].

As Devanagari characters, Marathi characters can be divided into three groups based on the presence and position of a vertical bar namely [5]:

1.1 Compound Character with End bar Character:

स्त ख्य ज्य
त्व म्य च्य
ल्प न्म न्न
त्प ल्य श्ल
घ्य त्य स्व

Fig. 1: End bar compound character.

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 5, Issue 1, pp: (33-39), Month: January - March 2017, Available at: www.researchpublish.com

1.2 Compound Character with Middle bar Character:

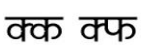

Fig. 2: Middle bar compound character.

1.3 Compound Character with No Bar character:

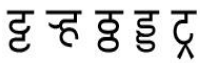

Fig. 3: No bar compound character.

There are many script and Languages in the world. The researchers have done work on some of them like English, Chinese, Latin, Arabic, Japanese, Thai and Devanagari. There are many pieces of work have been done towards the recognition of printed Devanagari Character and at present OCR systems are commercially available for some of the printed Indian scripts [6].

The research work on character recognition of Devanagari script was started in 1970 [4]. Compound character are those where two or more character are joined together to produce a special character. These characters are such type in which one half of character is connected to full character [4].

Half Symbol(s) + full symbol are called jodakshar.

In Marathi there is various form of compound character [5, 7]. For each compound character, it is corresponding to its "half" predefined character [5]. Some form of the compound character is shown below [5]:

i. Small cross line connected to the central vertical line of the first character

E.g क + र = क्र but it is क्र.

ii. If the first character does not have any vertical then the compound character will have another form :

E.g ट + र = ट्र but it is ट्र.

iii. If the first character is र then we have another form:

E.g र + ह = र्ह but it is र्ह

II. LITERATURE REVIEW

There are many script and languages in the in the world. The researchers have done work on some of them like English, Chinese, Latin, Arabic, Japanese, Thai and Devanagari [5]. Devanagari is the most widely used script in India. Sanskrit, Nepali, Hindi and Marathi are the Devanagari script used by more than 400 million people [8]. Unconstrained Devanagari writing is more complex than English language due to the possible variations in the shape, number and direction of the constituent strokes [8].

First research report on handwritten Devanagari characters was publishing in 1997[8]. At present researchers have started to work on handwritten Devanagari characters and few research reports are publish recently [8].

K. K. Kale et al proposed a Zernike moment feature extraction for Handwritten Devanagari compound character recognition and obtained result up to 98.37% and 95.82%. S. Shelke et al worked on Neural Networks for recognition of handwritten Marathi compound character and obtained 97.95% accuracy. P. E. Ajmire et al worked on seventh central moment for handwritten Devanagari compound character recognition and obtained 93.87% accuracy. Mrs. S. Golait et al worked on minutiae detection algorithm for handwritten Marathi compound character segmentation and obtained 90% accuracy. S. Arora et al proposed multilayer perceptron (MLP) for handwritten Devanagari character recognition and get 98.03% accuracy. K. V. Kale et al proposed SVM for handwritten and printed Devanagari compound character recognition and obtained the RR for printed is 98.42%(Basic) and 98.31%(compound) and for handwritten is 98.51%(Basic) and 98.30%(Compound).

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 5, Issue 1, pp: (33-39), Month: January - March 2017, Available at: www.researchpublish.com

III. PROPERTIES OF DEVANAGARI BASIC AND COMPOUND CHARACTER

A basic set of symbols of Devanagari script consists of 12 vowels, 36 consonants [9]. The alphabet of modern Devanagari script consists of 14 vowels and 33 consonants also called as basic characters [10]. The concept of upper/lower case is absent in Devanagari script [11]. In this script vowel following by a consonant takes a modified shape, these modified shapes are called modified characters [9, 10]. Occurrence of compound character in the Devanagari script is up to 8 to 10% [11]. A consonant or vowel following a consonant sometime takes a compound orthographic shape which we called as a compound character [9, 10]. Compound character can be combination of two consonants as well as a consonant and a vowel [9]. There are about 280 compound characters in Devanagari [11]. There are two routes for recognition of compound character. One way by separation of the character and second is without separation [11]. For first method two separate features are extracted and then recognition is done. For second method without separation the feature are extracted [11]. At present no dataset on Devanagari handwritten compound character is available.

IV. TEMPLATE MATCHING TECHNIQUE

In the proposed system, we aim at recognizing handwritten compound characters in Devanagari [12]. This is done by employing template matching technique [12]. The figure shows the block diagram of the proposed system, which consist of different phases starting with character input, pre-processing, structural classification, resized character and to generate a template and then matched with the similar templates in the dataset [12].

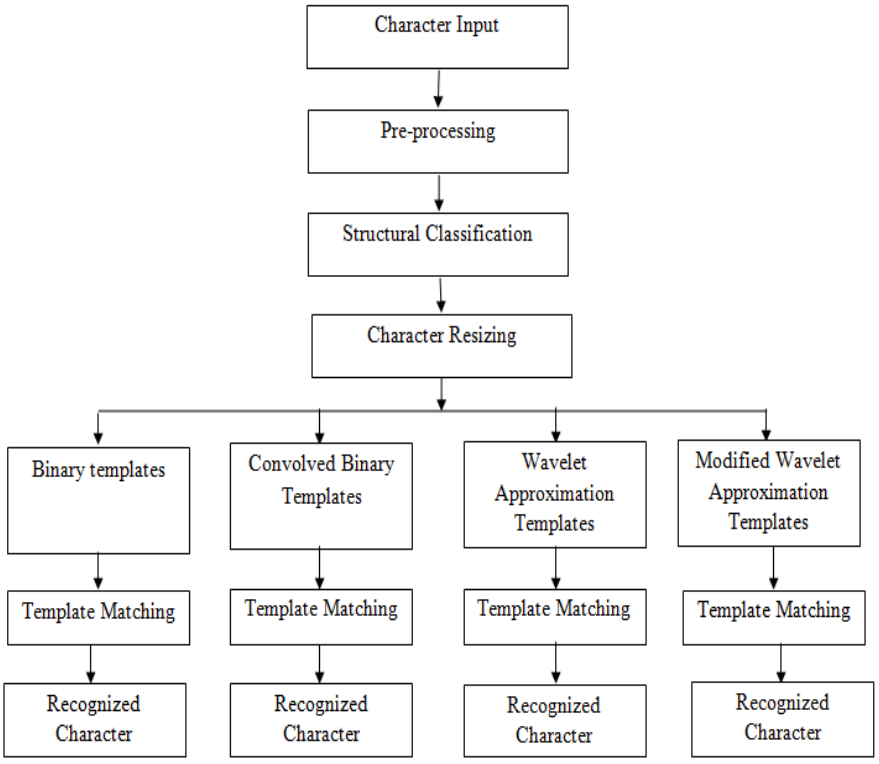

Fig 4. Template Matching Using Different Template

V. PREPROCESSING

Preprocessing is important in handwritten compound character recognition. As in handwritten character, some writer may put more trace on joining of characters or some may neglect the join. This means some character may have discontinuity in compound character. Preprocessing is necessary to remove this discontinuity in compound character.

The preprocessing plays an important role in handwritten compound character recognition [11]. Preprocessing is an important step by applying a number of procedures like smoothing, enhancing, filtering etc, for making digital image usable by subsequent algorithm to improve their readability [11]. The digital scanned image is first binaries, preprocess using various morphological operations opening closing, spurs operations which connect discontinues [11].

Pre-processing aims to produce data that are easy for the character recognition system to operate accurately [13]. The main objective of pre-processing are- 1) noise reduction; 2) normalization of the data; 3) compression in the amount of information to retained [13]. Many noise reduction techniques are three major groups [13]. The techniques are given in Table 1.

Table 1: Noise Reduction Techniques

Sr. no.	Category	Operation	Comments
1.	Filtering	Removes by noise and diminish spurious points, usually introduced by uneven writing surface and/or poor sampling rate of the data acquisition device [13].	Filters can be designed for smoothing, sharpening, thresholding, removing slightly texture or colored background and contrast adjustment purpose [13].
2.	Morphological	Filter the document image replacing by the logical operations [13].	Morphological operations can be successfully used to remove the noise on the document images due to low quality of paper ink, as well as erratic hand movement [13].
3.	Noise Modeling	Noise could be removed by some calibration techniques if a model for it were available [13].	There is little work on modeling the noise introduced by optical distortion, such as speckle, skew, and blurs [13].

A. *RGB to Gray Image:*

RGB image is three matrices which makes task more complex, so it is necessary to convert the image into gray. The gray images have values from 0 to 255.

In preprocessing the character images are converted to binary images using rgb2gray utility in MATLAB [9].

B. *Binarization:*

Binarization is the process of which converts colored image or gray scale image into binary form (0 & 1) by thresholding [14]. The method name thresholding is used to convert gray or colored image into binary image [14]. The binarization process is done by local or global thresholding [15]. Local thresholding methods are based on applying different threshold values to different regions of the image [15]. Global thresholding methods apply one threshold value to entire image [15].

C. *Filtering:*

To remove the noise present in the binaries image filtering has been done. Filtering also illustrated in above table [1].

D. *Boundary Tracing:*

Tracing of the boundary identifies the connected components of the characters in the filtered images and stores it in array [9].

E. *Normalization:*

Normalization is the process of converting the image into a standard size [14, 15]. In this, size normalization avoids inter class variation among characters [15]. Bilinear, Bi-cubic interpolation techniques are a few methods for size normalization [15, 16].

F. *Skeletonization:*

In Skeletonization, the thickness of the character is reduced to one-pixel character bound. Following figure shows the steps in preprocessing of images which are described above [9].

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 5, Issue 1, pp: (33-39), Month: January - March 2017, Available at: www.researchpublish.com

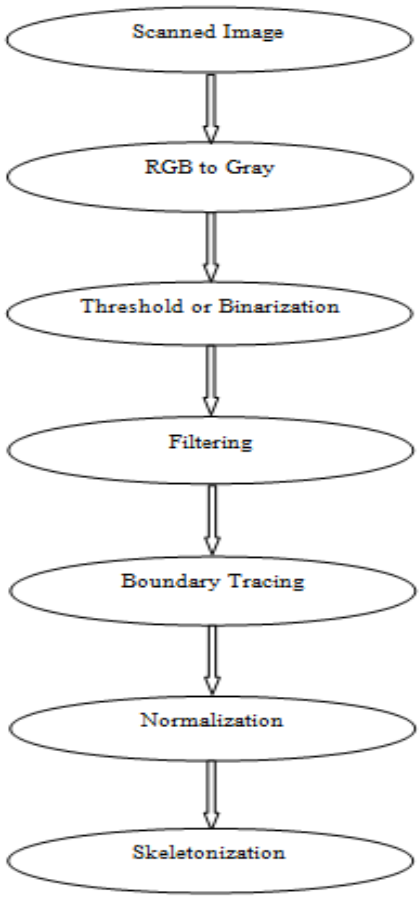

Fig. 5: Steps in Preprocessing of Image

VI. STRUCTURAL CLASSIFICATION

The large number of compound character set with a wide range of variations in the writing style demand a preclassification of the characters before the final recognition [2]. The pre-processing classification is done using a two stage classification based upon the structural features [2]. The two stage structural classification is followed by character resizing. The character is resizing to a typical size of 16*16 or 32*32 [12].

VII. RECOGNITION TECHNIQUE

In this system, the characters are recognized using template matching [12]. This is the simplest way of character recognition, based on matching the stored prototypes against the character or word to be recognized [17, 18, 19]. The matching operation determines the degree of similarity between two -vectors [18]. The basic template matching performs cross correlation of the 2D function f (m, n) with the template, g (m, n) [20]. The result contains peaks at the location of the matches between the template and underlying object [20]. The peak of cross correlation function (r) indicates the amount of matching as shown in following equation [20]. As a cross correlation function moves from 0 to 1, where higher the value, more is the similarity [12].

$$r = \frac{\sum_m \sum_n (f - \bar{f})(g - \bar{g})}{\sqrt{\left(\sum_m \sum_n (f - \bar{f})^2\right)\left(\sum_m \sum_n (g - \bar{g})^2\right)}} \quad \text{(eq……1)}$$

Where $\bar{f}$ and $\bar{g}$ are the mean of 2D function and the template [12]. A template matcher can combine multiple information sources, including match strength and k- nearest neighbor measurements from different matrices [17, 18, 19]. The recognition rate of this method is very sensitive to noise and image deformation [17, 18, 19].

Table 2: Template Matching Performance for Compound Character Recognition [12].

Sr. no.	Template generation technique	Resize factor	Template size	Recognition time for EBNE/11 (sec)	Recognition rate (%)
1.	Binary template	16x16	16x16	4.15	87.65
		32x32	32x32	5.79	89.00
2.	Convolved binary template	16x16	31x31	5.04	91.99
		32x32	63x63	5.68	92.41
3.	Wavelet approximation template	16x16	8x8	5.29	93.89
		32x32	16x16	5.44	94.39
4.	Modified wavelet approximation template	16x16	15x15	5.97	95.00
		32x32	31x31	6.22	95.50

VIII. CONCLUSION

Character recognition is one of the vital tasks in pattern recognition. The popularity of character recognition is increasing day by day with the advent of new, fast and efficient hardware and software.

India is multi-lingual country; so many more efficient and real-time text recognizers are required. A good text recognizer has many commercial and practical applications. Hence there is a need to develop a very good character recognition system which must achieve highest accuracy.

REFERENCES

[1] S. Arora, D Bhattacharjee, M. Nasipuri, D. K. Basu, M. K. Kundu, "Application of Statistical Features in Handwritten Devanagari Character Recognition", Internal Journal of Recent Trends in Engineering, vol. 2, no. 2, November 2009.

[2] S. Shelke, S. Apte ," A Multistage Handwritten Marathi Compound Character Recognition Scheme using Neural Networks and Wavelet Features" , International Journal of single processing, Image processing and Pattern Recognition, vol. 4, No. 1, March 2011.

[3] S. Shelke, Shaila Apte, "Multistage Handwritten Marathi Compound Character Recognition using Neural Networks", Journal of PATTERN RECOGNITION RESEARCH 2(2011) 253-268 Received January 15, 2011, Accepted August 2, 2011.

[4] S. V. Chavan, K. V. Kale, M. M. Kazi, Y. S. Rode, "Recognition of Handwritten Devanagari Compound Character a Moment Feature Base Approach", International Journal of Machine Intelligence, Volume 5, Issue 1, pp-421-425,2013

[5] P. E. Ajmire, R. V. Dharaskar, V. M. Thakare, " Handwritten Devanagari (Marathi) Compound Character Recognition using Seventh Central Moment", International Journal of Innovative Research in Computer and Communication Engineering ,vol. 3, Issue 6, June 2015.

[6] P. E. Ajmire, R. V. Dharaskar, V. M. Thakare, "A Comparative Study of Handwritten Marathi Character Recognition", International Journal of Computer Application, 2012.

[7] P. E. Ajmire, R. V. Dharaskar, V. M. Thakare, "Feature Extraction and Selection for Handwritten Devanagari Vowels Recognition", IJRET: International Journal of Research in Engineering and Technology, 2016.

[8] S. S. Golait, Dr. Latesh Malik, "Handwritten Marathi Compound Character Segmentation using Minutiae Detection Algorithm", International Conference on Computational Science, Precede computer science 87, pp 18-24, 2016.

[9] K.V. Kale, P. D. Deshmukh, S. V. Chavan, M. M. Kazi, Y. S. Rode ,"Zernike Moment Feature Extraction for Handwritten Devanagari (Marathi) Compound Character Recognition", (IJARAI) International Journal of Advanced Research in Artificial Intelligence, vol. 3, 2014.

[10] K.V. Kale, P. D. Deshmukh, S. V. Chavan, M. M. Kazi, Yogesh Rode, "Zernike Moment Feature Extraction for Handwritten Devanagari Compound Character Recognition ", Science and Information Conference 2013, pp-7-9, October 2013.

[11] K. V. Kale, S. V. Chavan , M. M. Kazi, Y. S. Rode ,"Handwritten and Printed Devanagari Compound using Multiclass SVM Classifier with Orthogonal moment Feature", International Journal of Computer Application (0975-8887), Volume 71, No. 24, June 2013.

[12] Compound character recognition using various models - Shodhaganga shodhganga. inflibnet. ac.in/ bitstream/10603/ 25484/16/16_chapter_08.pdf.

[13] A. S. Wanchoo, P. Yadav, A. Ansuse ," A Survey on Devanagari Character Recognition for Indian Postal System Automation", International Journal of Applied Engineering Research ISSN 0973-4562 volume 11, pp 4529-4536,2016.

[14] P. Verma, Pratibha Singh ," A Survey on Handwritten Character Recognition using Artificial Neural Networks", International Journal of Engineering Applied Sciences and Technology, vol. 1, issue 5, pages 80-86, march-April 2016

[15] Anitha M. Chacko, Dhanya P. M., " Handwritten Character Recognition in Malayalam Scripts- A Review", International Journal of Artificial Intelligence & Applications (IJAIA), vol. 5, No. 1, January 2014.

[16] Prof. Bamb, Kalpesh K., "A Literature Survey on Character Recognition of Indian Scripts for New Researchers", International Journal of Modern Trends in Engineering and Research, Date 28-30 April, 2016.

[17] C. Alpa, N. Mankad, N. Gondalia, "Hindi Character Recognition using Neural Network and Various Optimizing Techniques: A Survey" IJSRD-International Journal for Scientific Research & Development\vol. 2, Issue 10,2014.

[18] Ms. M. Shalini, Dr. B. Indira," Automatic Character Recognition of Indian Languages-A brief Survey", IJISET-International Journal of Innovative Science, Engineering & Technology, vol. 1, Issue 2, April 2014.

[19] V. J. Dongare , Vijay H. Mankar ," A Review of Research on Devanagari Character Recognition", International Journal of Computer Applications (0975-8887) volume 12-No. 2, November 2010.

[20] Fareeb Talib, Prof. Dipti Patil, "Multistage Classification for Handwritten Marathi Compound Letters using Wavelet Technology", 2012.