Handwritten Devanagari Compound Character Recognition using Legendre Moment an Artificial Neural Network Approach

[*1]K.V.Kale, IEEE Senior Member, [2]S.V.Chavan, IEEE Student Member,
[1]M.M.Kazi, IEEE Student Member, [1]Y.S.Rode, IEEE Student Member
[1]Dept. of Computer Science & I.T., Dr. Babasaheb Ambedkar Marathwada Universitiy, Aurangabad (MS) - India
[2]Dept. of Computer Science & I.T. MSS's Arts, Commerce & Science College, Ambad Dist. Jalna (MS) - India
kvkale91@gmail.com, shripc@gmail.com, mazhar940@gmail.com, ys.rode@gmail.com

***Abstract*— Handwritten Devanagari Compound character recognition is one of the new challenging task for the researcher, because Compound character are complex in structure, they are written by combination two or more character. Their occurrence in the script is upto 12 to 15%. In this research paper, a recognition system for handwritten Devanagari Compound Character is proposed bases on Legendre moment feature descriptor are used to recognize. Moment function have been successfully applied to many pattern recognition problem, due to this they tends to capture global features which makes them well suited as feature descriptor. The process image is normalized to 30X30 pixel size divided into zone, from this structural as well as statistical feature are extracted from each zone. The proposed system is trained and tested on 27000 handwritten collected from different people. For classification we have used Artificial Neural Network. The overall recognition rate for basic is upto 98.25% and for all compound character is 98.36%.**

Keywords- Handwritten Devanagari Compound; Legendre Moment; ANN; Pattern Recognition;

I. INTRODUCTION

Handwritten character recognition has always been a challenging task in image processing and pattern recognition. There are five major stages in the handwritten character recognition (HCR) problem: Image preprocessing, feature extraction, training, testing and post processing. The feature extraction method is probably the most important stage in achieving high recognition performance. This field of research is applicable for application areas i.e. form filling, job application, bank and postal automation [1-3] and also an identification of a person of a scanned handwritten script; it is useful for biometric modality with application in forensic and historic document analysis (HDA) and represents an excellent study area within the research field of biometrics. In the proposed system, we are extracting Legendre moment features descriptor, from the scanned images of handwritten. For recognition purpose we have used Artificial Neural Network.

Recognition of Handwritten in Indian script [4] is also one of challenging task specially, for several reasons because of complex structure of character with their modifiers and presence of compound character. Writing style in Devanagari script is from left to right. The concept of upper/lower case is absent in Devanagari script. Compound characters are those where one half of character is connected to full character to produce a special character. Thus there are large variations in shape of character as writing style, pen quality (thick/thin), strokes that substantial extent the recognition accuracy.

Work on Devanagari was started in 1970. Sinha and Mahabala[5] presented a syntactic pattern analysis system for the recognition of Devanagari characters (DC). OCR work for printed and handwritten characters in various Indian scripts [6-8] is carried out by researchers but major work is found for Bangla[9,10]&Devanagari. First research report on handwritten DC was published in 1977 by Sethi and Chatterjee[11]. An extensive research work on printed Devanagari was carried out by Bansal et.al. [12-14]. Work on Handwritten Numerals of Devanagari is carried out by researcher is presented in [15-16]. Central, invariant, Zernike moments use for recognition of character of different languages reported in [17-25]. In [25-27] research has proposed Chain Code Histogram and directional information obtained from the arc tangent of the gradient for feature extraction. A significant contribution is due to Arora et al. in which they proposed a multi-feature extraction based technique in [28]. Pal et al. proposed SVM and MQDF based scheme for recognition of Devanagari [29]. U. Pal and T. Wakabayashi [30] proposed a comparative study of different DC recognizers using features based on curvature and gradient information. Sushama Shelke et al. [31] presented a novel approach for recognition of unconstrained handwritten Marathi characters. Baheti M.J. et. al [32] proposed a method based on AIM for Gujarati numerals uses KNN and PCA as classifiers. Recognition of handwritten Bangla compound character was attempted by U. Pal et al. [33] using gradient features. Work on handwritten Marathi compound characters found in [34-35] by S. Shelke and S. Apte. From the above discussion it is evidence that moment can be considered as potential features for recognition of character and numerals, which motivate us to enrich the several orthogonal and discrete moment features and test the efficacy of the system for compound characters. While significant advances have been achieved in recognizing Roman-based scripts like English, ideographic characters Chinese, Japanese, Korean, and Arabic to some extent. Only few works on some of the major Indian scripts like Devanagari, Bangla, Gurumukhi, Tamil, Telugu, are available in the literature.

The paper is organized as follows: Section 2 deals with Moments discussion. Database collection & Preprocessing for the experiment in section 3. Section 4 presents Feature extraction procedure. Section 5 details of the classifier used for recognition. The experimental results are discussed in

978-0-7695-5066-4/13 $26.00 © 2013 IEEE
DOI 10.1109/ISCBI.2013.62

Section 6. Finally, conclusion on the paper is given in Section 7.

II. MOMENT

The concept of moment in mathematics evolved from the concept of moment in physics. It is an integrated theory system. For both contour and region of a shape, one can use moment's theory to analyze the object.

Moment based techniques have been successfully applied to several image processing problems and they represent a fundamental tool for generating feature descriptors. Feature descriptors built from moment functions capture global features, and as such they are well suited for shape and character recognition. Some moment functions exhibit natural invariance properties, such as invariance to rotation, translation or scaling. Translation and scale invariance is usually obtained by normalizing the input image with its geometric moments. The original image can be reconstructed from its moments, as an infinite series of moments uniquely identify a specific distribution. Reconstruction from orthogonal moment functions can be done.

A. Legendre moments (LM)

The basis function for Legendre moment is $\phi_{nm}(x, y) = P_n(x)P_m(y)$ where $P_p(x)$ denotes the pth order of Legendre polynomial. The $(n+m)^{th}$ order of Legendre moment, L_{nm}, is defined in Eq. 1 [38].

$$L_{nm} = \frac{(2n+1)(2m+1)}{4} \int_{-1}^{1}\int_{-1}^{1} P_n(x)P_m(y)f(x,y)dxdy.$$

where $P_n(x)$ is the n^{th} order of Legendre polynomial give by

$$P_n(x) = \frac{1}{2^n} \sum_{k=0}^{n/2} (-1)^k \frac{(2n-2k)!}{k!(n-k)!(n-2k)!} x^{n-2k} \text{ n} \quad (2)$$

Since the Legendre polynomials are orthogonal over the interval [−1, 1], the image f(x, y) can be reconstructed from its moments. Teague derived a simple approximation to the inverse transform for a set of moments through order M given by

$$f(x,y) \approx \sum_{n=0}^{M} \sum_{m=0}^{n} L_{n-m,m} P_{n-m}(x) P_m(y) \quad (3)$$

III. DATA COLLECTION & PREPROCESSING

The basic set of symbols of Devanagari script consists of 12 vowels (or swar), 36 consonants (or vyanjan) as shown in Fig. 1, which are used for this work, we have also used 45 compound character + 15 split component of compound character for proposed research which is presented in Fig 2 and Fig 3.

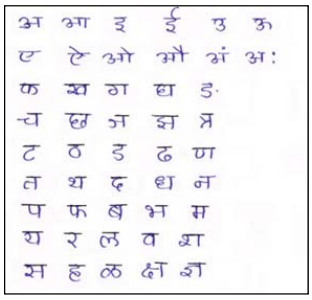


Fig.1-Sample dataset of handwritten Devanagari Character

Compound characters can be combinations of two consonants as well as a consonant and a vowel. Compounding of three or four characters also exists in the script. The compound characters are joined in various ways, by removing vertical line of the character and then join it to the other character from left side like  , or another way is join side by side or one above the other  . The e.g. regarding to the compound is shown in Fig.(2).Split character is the half character of basic character which get connected to other character e.g. split is given in Fig.(3).

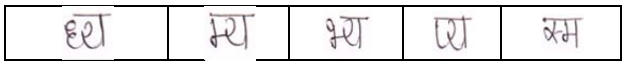


Fig.2 Example of Compound Character

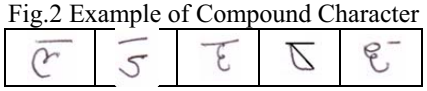


Fig. (3) Example of Split of Compound Character

A. Database Designing

For the proposed work, the database is prepared by scanning the images of Handwritten Devanagari Basic & Compound Character, which are collected on a special designed sheet, from individuals of various professions. Source document are scanned at 300 dpi using canon flatbed scanner and stored in TIFF format images in 1 byte per pixel, and stored for extraction of isolated character. All output samples were checked manually and necessary editing/removal were done. Image samples of the present databases are also maintained in the original form, the writer other information is also available for the entire document.

At present no dataset on Devanagari handwritten compound characters is available. A significant contribution of present work is the pioneering development of large database for Handwritten Devanagari Basic and Compound Character was collected. Details of this database are provided in Table I. Handwritten Devanagari (Basic & Compound) character consists of 27000 samples written by writer from different location, fields, profession.

TABLE I. DATASET OF HANDWRITTEN DEVANAGARI BASIC AND COMPOUND CHARACTER

Script	Character Set	Basic Character	Comp-ound Character	Split Comp-onent of Compound Character	Total
Devanagari	Training	9600	9000	3000	
	Testing	2400	2250	750	

Character	Total	12000	11250	3750	27000

B. Preprocessing

The preprocessing plays important role in handwritten character recognition as in pattern recognition task. Preprocessing is an important step of applying a number of procedures for smoothing, enhancing, filtering for making a digital image usable by subsequent algorithm in order to improve their readability. The morphological opening and closing operators not only remove image noise but also connect discontinuities. The next step is skeletonization which reduce the pen width and computation cost.

To achieve the recognition accuracy of Devanagari Compound character the structural classification is necessary. There are two methods to recognized compound character, the one is partitioned of the character and other is without partition of character. For the first one, two separate features are extracted and then recognition is done and for second method without separation the features are extracted.

C. Pre-classification

With the help of above processing step, the pre-classification of character is done which is based on two stages global features i.e. presence of vertical line, position of it in the character and enclosed region in the character and second is local features i.e. end points and junction in the character. On the basis of global feature, the character is classified into three major categories, presence of vertical bar, at right character and at middle, and absence of vertical bar.

Vertical bar on right are further classified into two categories based on whether the vertical bar and rest of the character are connected or not to the bar. These are show in the following Table II.

TABLE II. CLASSIFICATION OF DEVANAGARI CHARACTER

Sr. No	Pre-classification	Character
1	Character connected with vertical bar at right side	Ká, Iá, Ûî, pe, Pá, $e, Ie, Le, Oá, ve, He, ye, Yá, cá, Ùe, ue, Já, mî,<e, #î, %e
2	Character not connected with vertical bar at right side	iá, Ce, Má
3	Character with vertical bar at middle	Jáâ, heâ
4	Character with absence of vertical bar	[.,Ú, š, ", [, {, o, j, n, U

D. Local Structural Classification

The local features are detected on the bases of the end points of the character. To detect the end points there are two steps, first partitioned the image into 3x3 i.e. 9 quadrants and secondly detect the end points and junction in the individual block as shown in Figure 4.

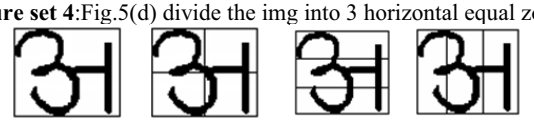
Fig 4: Presence of End Points in partition block of Character

In this paper basic, compound character are written on a plain paper. The characters are scanned, pre-processed by the above discussion and store automatically for database creation. The character is processed, normalized into the required size and transfer for structural classification and further features are extracted which are used for training and testing.

IV. FEATURE EXTRACTION

Moment based features are extracted from the each zone of the scaled character bitmapped image. The image is partitioned into zone and features are extracted from each zone as shown in Fig 5. In this paper Legendre moments based feature extraction is proposed for off-line Devanagari Handwritten Basic and Compound Character. To get the feature set, at first, the image is segmented to 30 x 30 blocks, and partitioned as feature set as follows.

Feature set 1: Fig 5(a) shows is considered as whole.
Feature set 2: Fig. 5(b) divide the img into four equal zones.
Feature set 3: Fig.5(c) divide the img into 3 vertical equal zones
Feature set 4:Fig.5(d) divide the img into 3 horizontal equal zones



a **b** **c** **d**

Fig.5 – Partition of Devanagari Character into feature set

TABLE III. : MOMENT FEATURE SET FOR DEVANAGARI

Zone set	Moment Feature	Feature Set
FS1	Legendre	9
FS2	Legendre	36
FS3	Legendre	27
FS4	Legendre	27

In pattern recognition, feature extraction stage in character recognition plays a major role in improving the recognition accuracy. Features are extracted from binary image/characters. Many characters are misclassified due to their similarity in shape or slight variation in writing style. So features which are selected should tackle these problems.

Block Diagram of Recognition System

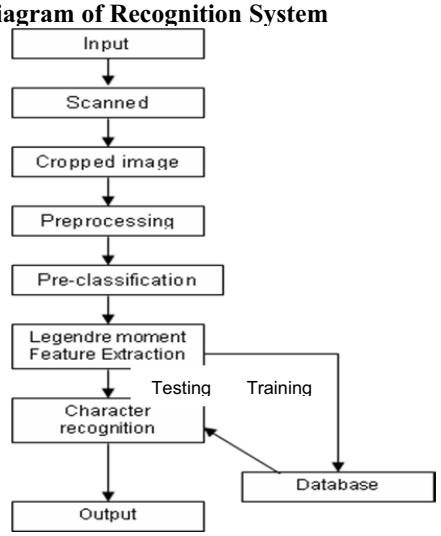


Fig 6: Block Diagram for proposed work

V. CLASSIFICATION & RECOGNITION

For the purpose of training and testing Artificial Neural Network is used for classification. The network is trained with a set of handwritten Devanagari Basic and Compound Character. The network accepts Legendre moment feature of each character in the form of feature set, 9, 36, 27, 27 as input. The experimental result obtains by using this feature to recognized Handwritten Devanagari Basic and Compound Character is presented in next section. At this stage the character goes through the pre-classification and then the Legendre moment features are extracted.

The selected architecture of the network is a two layer feed forward neural network with 108 neuron each. The transfer function is log-sig. The network is trained with Gradient Decent Backpropagation with adaptive learning rate. The learning rate and momentum term are set to 0.7 and 0.5 respectively. The performance function uses sum square error. The goal was set to 0.1 The network is trained with moment feature until the sum squared error is 0.1 After training the network is ready to use for testing.

VI. RESULT AND DISCUSSION

The proposed system deals with the Handwritten Devanagari basic and compound character the preparation of training and testing set are as discussed in Section 3.1. Results are shown in Table 2. The database consists of 12000 basic Devanagari characters from which we have selected 9600 for training and 2400 for testing. The database also consists of 11250 compound and 3750 split Devanagari Character First structural pre-classification is done as discuss in Section 3.2 and then moment based features are extracted from feature set as discuss in Section 4. The results are encouraging and average recognition accuracy is 98.25% is obtained.

TABLE IV. : RECOGNITION RATE OF SAMPLE DEVANAGARI BASIC CHARACTER IN %

Devanagari Char	Feature Set			
	FS1	*FS2*	*FS3*	*FS4*
क	90.24	98.75	98.46	98.52
ख	90.25	98.86	98.55	98.54
ग	91.27	98.87	98.64	98.62
ल	90.54	98.85	98.57	98.56
ट	91.25	98.75	98.71	98.52

From the sample of experiment conducted it can be seen that the character ie has the average recognition rate i.e. 96.85 so the highest recognition rate i.e. 98.87% for feature set FS2, and lower rate is 91.27% for Feature set 1.

Secondly from the selected sample the recognition rate for ट we have got average recognition rates upto 96.81% from the feature set the maximum recognition rate for it is 98.75% and minimum is 91.25%. From the feature set FS2 has a highest recognition rate for all character as compared to other FS1, FS3 and FS4. Thus overall recognition rate for 48 Devanagari Basic Character is 98.25%.

TABLE V. :RECOGNITION RATE OF SAMPLE DEVANAGARI COMPOUND CHARACTER IN %

Devanagari Chara	Feature Set			
	FS1	*FS2*	*FS3*	*FS4*
क्य	94.25	98.66	97.57	97.45
त्य	94.85	98.76	97.86	98.55
प्र	93.55	98.72	97.52	98.54
न्य	93.68	98.85	97.64	98.65
फ्य	94.88	98.75	97.85	98.74

Recognition performance for sample of handwritten compound character is shown in Table 4. In this, the letter फ्य have got average recognition rate as 97.56% and from feature set FS2 has maximum 98.75% and minimum as 94.88% from FS1. Next the character त्य have highest average recognition rate i.e. 97.51% and from feature set FS2 has given maximum 98.76% and minimum 98.55%. The feature set FS2 has gives overall better result as compared to FS1, FS2 and FS3. Thus it can be conclude that the recognition rate for all compound character is 98.36%.

VII. CONCLUSION

In this paper we have proposed handwritten Devanagari Compound character recognition system from which Legendre moment based features extracted and using simple Feed Forward Neural Network the system is trained and tested. We have obtained better result from the Legendre moment based feature. It gives better result for basic and compound character. As the image are partitioned in feature set FS1,FS2,FS3 and FS4, the feature set FS2 have given better result to all character as compared to other reference. Further to get better result we are going to modify the system with other orthogonal moment features set.

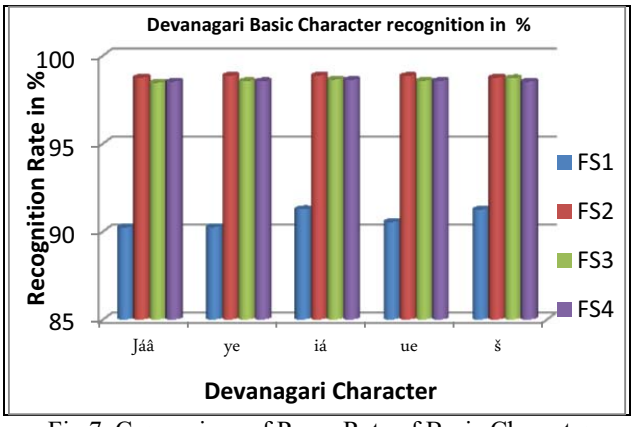


Fig 7. Comparison of Reco. Rate of Basic Character

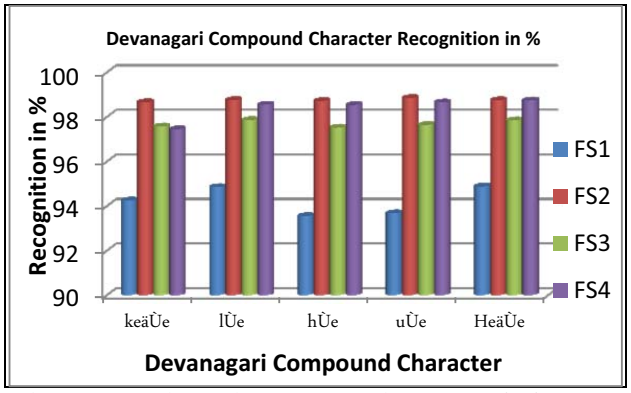


Fig 8. Comparison of Reco. Rate of Compound Character

Reference

[1] K. Roy, S. Vaidya, U. Pal, B. B. Chaudhuri, A. Belaid, "A System for Indian Postal Automation", Proc. 8th Int. Conf. Document Analysis and Recognition, Seoul, Korea, Aug. 31-Sep. 1, 2005, pp.1060-1064.

[2] U. Pal,R. K. Roy, K. Roy, F. Kimura, "Indian Multi Script Full Pin-code String Recognition for Postal Automation", Proc. 10th Int. Conf. Document Analysis and Recognition, Barcelona, Spain, Jul. 26-29, 2009, pp. 456-460.

[3] B. B. Chaudhari, Digital Document Processing - Major Directions and Recent Advances, London:Springer, 2007.

[4] U. Pal and B. B. Chaudhari, "Indian Script Character Recognition: a Survey", Pattern Recognition,vol. 37, 2004, pp. 1887-1899.

[5] Sinha R.K., Mahabala 1979 "Machine Recognition of Devnagari Script", IEEE Trans. System Man Cyber Pgs 435-441

[6] B. B. Chaudhuri and U. Pal, "A complete printed Bangla OCR system", Pattern Recognition, vol. 31, 1998, pp. 531-549.

[7] K. Roy, T. Pal, U. Pal, and F. Kimura, "Oriya Handwritten Numeral Recognition System", Proc. 8th Int. Conf. Document Analysis and Recognition, vol. 2, Seoul, Korea, Aug. 31-Sep. 1, 2005, pp. 770 – 774.

[8] Arun Pujari, C. Dhanunjaya Naidu, M. Sreenivasa Rao, and B. C. Jinaga, "An Intelligent Character Recognizer for Telugu scripts using Multi resolution Analysis and Associative Memory", Image and Vision Computing, vol. 22, 2004, pp. 1221-1227.

[9] U. Bhattacharya, S. K. Ghosh, and S. K. Parui, "A Two Stage Recognition Scheme for Handwritten Tamil Characters", Proc. 9th Int. Conf. Document Analysis and Recognition", Parana, Sept. 23-26, 2007, pp. 511 – 515.

[10] B. B. Chaudhuri, and U. Pal, "A Complete Printed Bangla OCR System", Pattern Recognition, vol. 31, no. 5, 1998, pp. 531-549.

[11] I.K. Sethi, and B. Chatterjee, Machine Recognition of Constrained Hand Printed Devnagari.Pattern Recognition, vol. 9, no. 2, 1977, pp. 69 – 75.

[12] V. Bansal, "Integrating Knowledge Sources in Devanagari Text Recognition", Ph.D. Thesis, IIT Kharagpur, 1999.

[13] V. Bansal, R.M.K. Sinha, "Partitioning and Searching Dictionary for Correction of Optically Read Devanagari Character Strings", Proc. 5th Int. Conf. Document Analysis and Recognition, Bangalore, India, Sept. 20-22, 1999, 6pp. 53-656.

[14] V. Bansal, R. M. K. Sinha, "On How to Describe Shapes of Devanagari Characters and Use them for Recognition", Proc. 5th Int. Conf. Document Analysis and Recognition, Bangalore, India, Sept. 20-22, 1999, pp. 410-413.

[15] Reena Bajaj, Lipika Dey, and S. Chaudhury, "Devnagari numeral recognition by combining decision of multiple connectionist classifiers", Sadhana, Vol.27, part. 1, 2002 pp.-59- 72.

[16] P.M. Patil, T. R. Sontakke, "Rotation, Scale and Translation Invariant Handwritten Devanagari Numeral Character Recognition using General Fuzzy Neural Network", Pattern Recognition, vol. 40, 2007, pp. 2110-2117.

[17] L.C.Barczak,M.J.Johnson,C.H.Messom, "Revisiting Moment Invariant: Rapid Feature Extraction and Classification for Handwritten Digit", Proceeding of Image and Vision Computing, Hamilton, New Zealand, December, 2007, pp.137 – 142.

[18] Duda, R.O.,Hart, P.E., Stork, D.G., Pattern Classification, Second ed. John Wiley and Sons, Inc. 14, 2001.

[19] Hamid Reza Boveiri, "Persian Printed Numeral Character Recognition using Geometrical Central Moments and Fuzzy Min Max Neural Network", International Journal of Signal Processing, 2009, pp. 226 – 232.

[20] Hamid Reza Boveiri, "Persian Printed Numeral Classification using Extended Moment Invariants", World Academy of Science, Engineering and Technology 63, 2010, pp. 167-174.

[21] S.Arora, D. Bhattacharjee, M. Nasipuri, D.K.Basu, M. Kundu, "Application of Statistical Features in Handwritten Devanagari Character Recognition", International Journal of Recent Trends in Engineering, Vol 2, No. 2, November 2009, pp. 40 – 42.

[22] R.Sanjeev Kunte and R.D.Sudhaker Samuel, "A Simple and efficient optical character recognition system for basic symbols in printed Kannada text", Sadhana, Vol. 32, Part 5, October 2007, pp.21-533.

[23] S.N.Nawaz et. al., "An approach to offline Arabic Character Recognition using Neural Network", IEEE ICECS, 2003, pp. 1325 – 1331.

[24] S. Kumar, and C. Singh, A Study of Zernike Moments and Its Use in Devnagari Handwritten Character Recognition. Proc. Intl. Conf. Cognition & Recognition, Mandya (India), 2005, pp. 514 – 520.

[25] N. Sharma, U. Pal, F. Kimura, and S. Pal, Recognition of Offline Handwritten Devnagari Characters Using Quadratic Classifier. Proc. Indian Conf. Computer Vision Graphics & Image Processing, Madurai (India), 2006, pp. 805 – 816.

[26] M. Hanmandlu, O.V. Ramana Murthy, and V.K. Madasu, Fuzzy Model Based Recognition of Handwritten Hindi Characters. Proc. Ninth Biennial Conf. Australian Pattern Recognition Society on Digital Image Computing Techniques & Applications, Glenelg (Australia), 2007, pp. 454 – 461.

[27] U. Pal, N. Sharma, T. Wakabayashi, and F. Kimura, Off-line Handwritten Character Recognition of Devnagari Script. Proc. Ninth Intl. Conf. Document Analysis & Recognition, Curitiba (Brazil), 2007, pp. 496 – 500.

[28] S. Arora, D. Bhattacharjee, M. Nasipuri, D.K. Basu, and M. Kundu, Combining Multiple Feature Extraction Techniques for Handwritten Devnagari Character ecognition. Proc. IEEE Region 10 Colloquium and Third Intl. Conf. Industrial & Information Systems, Kharagpur (India), 2008.

[29] U. Pal, S. Chanda, T. Wakabayashi, and F. Kimura, Accuracy Improvement of Devnagari Character Recognition Combining SVM and MQDF. Proc. Eleventh Intl. Conf. Frontiers in Handwriting Recognition, Montreal (Canada), 2008 pp. 367 – 372.

[30] U. Pal, T. Wakabayashi, and F. Kimura, Comparative Study of Devnagari Handwritten Character Recognition Using Different Feature and Classifiers. Proc. Tenth Intl. Conf. Document Analysis & Recognition, Barcelona (Spain), 2009, pp. 1111 – 1115.

[31] S. Shelke, S. Apte, —A Novel Multi-feature Multi-Classifier Scheme for Unconstrained Handwritten Devanagari Character Recognition, 12th International Conference on Frontiers in Handwriting Recognition, 2010.

[32] Baheti M. J., Kale K.V. Jadhav M.E. "Comparison Of Classifiers For Gujarati Numeral Recognition", International Journal of Machine Intelligence (IJMI), Volume 3, Issue 3, 2011, pp-160-163.

[33] U. Pal, T. Wakabayashi, and F. Kimura, "Handwritten Bangla Compound Character Recognition using Gradient Feature", Proc. 10th Int. Conf. Information Technology, Orissa, India, Dec. 17-20,2007, pp. 208-213.

[34] S. Shelke, S. Apte, "A NovelMultistage Classification andWavelet Based Kernel Generation For Handwritten Marathi Compound Character Recognition", Proc. Int. Conf. Communications and Signal Processing, Kerala, India, Feb. 10-12, 2011, pp. 193-197.

[35] S. Shelke, S. Apte, "A Multistage Handwritten Marathi Compound Character Recognition Scheme using Neural Networks and Wavelet Features", International Journal of Signal Processing, Image Processing and Pattern Recognition, vol. 4, no. 1, 2011, pp. 81-94.

[36] S. Saharia, P. K. Bora, and D. K. Saikia, .A comparative study on discrete orthonormal Chebyshev moments and Legendre moments for representation of printed characters,. in Proc. 4th ICVGIP, 2004, pp.491.496.

[37] Prokop RJ, Reeves AP. A survey of moment-based techniques for unoccluded object representation and recognition. CVGIP: Graphical Models and Image Process. 1992;54(5):438–460.

[38] Teague MR. Image analysis via the general theory of moments. J Opt Soc Am. 1980;70:920–930.