A Survey on Offline Recognition of Handwritten Devanagari Script

Ashwin S Ramteke, Milind E Rane

Abstract - Devanagari Script is mostly useful in India for writing number of official forms and documents, especially in the banking applications for writing amount in words on the cheque. The Offline recognition of handwritten Devanagari script has great application in automatic processing of handwritten bank cheques images, documentation of various official documents as well as digitization of government and non government documents. In this paper it has been tried to achieve automatic recognition of handwritten Devanagari Script by using various algorithms.

Keywords— Handwritten Devanagari Script, Preprocessing, Feature extraction, Classifiers.

———————————— ◆ ————————————

I. INTRODUCTION

Handwriting is the most natural mode of collecting, storing, and transmitting information which also serves not only for communication among human but also serves for communication of humans and machines. Handwritten character/script recognition is one of the interesting field based on the image processing and pattern recognition application.

Handwritten script recognition finds application in automation in various fields like postal automation [1,2], bank automation [3,4,6,8,], form filling , the automatic processing of bulk amount of papers, transferring data into machines, and web interface to paper documents [9].

Handwritten Indian script (Devanagari) recognition [5] is tough task due to the various characteristics of these scripts like their large Character set, complex shape, presence of modifiers, presence of compound characters and similarity between characters.

Devanagari script is mostly useful for writing and documentation purpose of most of the Indian languages such as Hindi, Marathi, Sindhi, Nepali, Sanskrit, Konkani, Maithili etc [11].

In this paper a survey is made in concern with this field of recognition of script/character and also some of the outcomes obtained till the date are also discuss herewith.

II. CHARACTERISTICS OF DEVANAGARI

One of the defining aspects of Indian script is the repertoire of sounds it has to support. As there is typically a letter for each of the phonemes in Indian languages, the alphabet set tends to be quite large. Devanagari originated from ancient Brahmi script through various transformations. India's official national language Hindi, written in the Devanagari alphabet has the most speakers which can be more than 500 million. The complex nature of the Devanagari script will be governed by the following characteristics.

Devanagari script has 13 vowels which are as shown in Figure 1, 34 consonants which are shown in fig.2, along with 14 modifiers of vowels and of "rakar," as shown symbols in Fig.3 respectively.
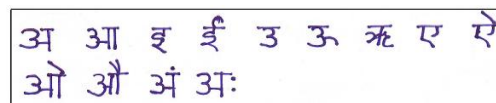

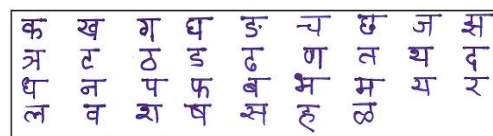Fig.1 Vowel in Devanagari

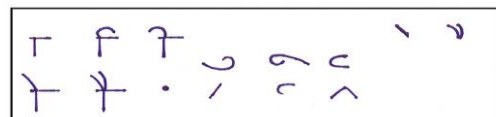Fig.2 Consonants in Devanagari

Fig.3 Modifiers in Devanagari

Also Devanagari having compound characters which are formed by combining two or more basic characters. To form the compound character in Devanagari half form of consonants are considered which are as shown in fig.4.
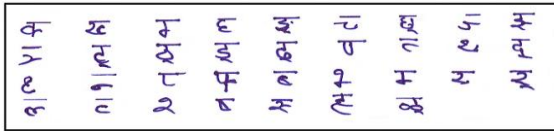
Fig.4 Half Forms of Consonants in Devanagari

There is no concept of upper or lowercase characters in Devanagari. It is a phonetic and syllabic script, words are written exactly as they are pronounced. Syllabic means that text is written using consonants and vowels that together form syllables. The vowels in can be either independent or dependent.

Apart from the above features another distinctive feature of Devanagari is the presence of a horizontal line on the top of all characters. This line is known as header line or "shirorekha" as shown in fig 5. The words can typically be divided into three strips: top, core, and bottom, as shown in Fig 5. When two or more characters appear side by side to form a word in Devanagari, the header lines touch and generate a bigger header line.

Fig. 5 Three Strips of Devanagari Word with Header Line

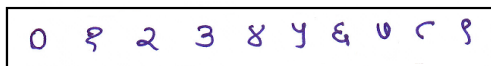Following figure 6 shows the Devanagari numerals.

Fig. 6 Devanagari Numerals

III. PROCESS OF RECOGNITION

The process of recognition of handwritten scripts is a tedious job. The process of recognition of handwritten script includes preprocessing and segmentation, feature extraction and recognition or classification processes which is totally based on the application of signal processing subjects like image processing, pattern recognition, artificial neural network.

Mostly all the features can be of two types, namely statistical and structural. The statistical features are derived from statistical distributions of points, such as moments, zoning, histograms, or projection features. Structural features are mainly based on geometrical properties of a symbol or character, like loops, directions of strokes, intersections of strokes, and end points.

Researchers mostly follow two approaches in handwritten character/script recognition. First is segmentation-based approach and the other is segmentation-free approach or holistic approach [14, 10, 17]. In the first approach, the words are initially segmented into characters or pseudo characters, and then, recognized. As a result, the success of the recognition module depends on the performance of the segmentation technique. The second approach treats the whole word as a single entity and it recognizes without doing explicit segmentation. Majority of the work in the recognition of handwritten Devanagari script has been done by Pal, Chaudhari and their team [1, 2, 4, 11, 18, 19] as well as by R.M.K. Sinha at IIT Kanpur [12, 21].

The recognition process includes following steps.

A. Preprocessing

Preprocessing is necessary when the data or image is capture for further processing. The noise is introduced in the image while acquisition of image or data, while transferring the data, as well as because of changing of some parameters of acquisition system in the optical character recognition. The another difficulty which introduced is the skew in the handwritten script/word. Skew is a distortion that is often introduced during scanning or copying of a document and it is unavoidable. Skew angle is the angle that text lines deviate from the x-axis. Since page decomposition techniques require properly aligned images as input, document skew must be corrected in advance; otherwise, serious performance degradations will result.

In general there can be three types of skew within a page, a global skew, when all text lines have the same orientation; multiple skews, when some text lines have a different orientation than the others; and non-uniform skew, when the orientation fluctuates within a text line. It must be noticed that a handwritten document image is usually expected to have multi-skew or even worse, non-uniform skew [16]. Once the global orientation is detected, the document skew can be corrected by a rotation at this angle. In other words, "skew correction" is applied after "skew detection". Another inherent phenomenon which is present because of writing way of a person, which known as slant. Skew is the deviation of the baseline from the horizontal direction and slant is deviation of average near-vertical strokes from the vertical direction. Following fig. 7 and fig. 8 is the example of skew and slant in the handwritten word respectively.

Fig.7 Skew in handwritten word

ज्ञावीस

Fig.8 Slant in handwritten word

Preprocessing techniques include noise removal, size scaling, and binarization of image, skew correction, slant correction, segmentation including the header line removal.

For preprocessing [46] first smoothed the image by using median filter and then binarized by using Nobuyuki Otsu thresholding method [7]. The binarized image is smoothed again using median filter. Mahesh Jangid [24] also used Otsu's method for binarization after adjusting the image intensity by imadjust () function in Matlab.

In [15] S.Arora and other followed initially Size Determination, where the approximate dimension of the character by forming a tight fit rectangular boundary around the character. Then the Distortion in the image is removed by thickening, thinning and pruning for removing distortions. The image is thickened first and then thinned to convergence. This gives a smooth one-pixel wide image of the character, which is pruned to remove the small projections resulting from the thinning algorithm. Finally normalization done by scaling to 100×100 pixels using affine transformation.

A 3x 3 kernels median filtering is applied followed by morphological smoothing to clean the spatial noise characterized by small components is used by M.Mehta, R.Sanchati and A.Marchya in [6]. It will eliminate the image abnormalities such as scanning error and ink variation. Haji [16] compared the three binarization algorithms in his MSC thesis. The Otsu's Method, the Liu and Srihari's Method, The Wu and Manmatha's Method. Normalizing , smoothing by using Gaussian filter, followed by interpolation to give a fixed number points is perform in the preprocessing techniques used by H. Swethalakshmi, A.Jayaraman, V.S. Chakravarthy, C.C. Sekhar in [26].

In implementation of this project for preprocessing, the image is captured by scanner with 96 dpi. , then it scaled using imresize ( ) function in matlab. The intensity of image is adjusted by using imadjust ( ). The noise in the image is recovered by using median filter approach. The binarization of image is perform by using the matlab function im2bw ( ).

To detect the shirorekha, it is assumed that the first pixel available for a given character, when looked from right to left, is the shirorekha, i.e. nothing can go past the shirorekha as is normal practice. This last pixel is traced out for the character and the shirorekha movement is detected by priority based neighborhood search. Differential distance based technique on the detected shirorekha

determines whether it is a near straight line or not. This technique is based on calculating the successive differences of the distances of upper envelop of the curve from the reference line. Reference line is top of the character image. Also, the shirorekha should terminate in an open end. If not, the shirorekha is rejected and the character is said to have no shirorekha [15]. In the [41] paper, a side histogram approach is used to remove the shirorekha. Where the minima in the histogram marks the separation of lines. In side histogram of the line the global peak in the top one-third portion is then taken to be the general position of the header line, and the portion above it is then cut off.

In implementation of this project for removal of shirorekha a simple pixel count algorithm is applied where the number of black/0 pixel in row are counted and if it is greater than observed threshold then they change to white/1.

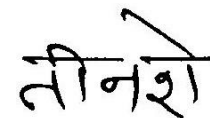Following fig. 9 and fig. 10 will show the input image and the output image with shirorekha removed respectively.

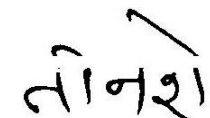तीनशे

Fig.9 Input Image with shirorekha

तीनशे

Fig.10 Output image without shirorekha

For skew angle detection, method proposed [22] in a skew correction scheme is designed for non-uniform skew, which assumes that the Matra is piecewise linear. Rajiv kapoor, Deepak bagai, T.S. kamal proposed a method for skew angle detection in [32]. This method is applicable to the whole document image. The method is based on Shirorekha as inherent feature of Devanagari. The method gives fairly good accuracy and this work stresses on the exploitation of the Shirorekha feature. A.Roy, T.K.Bhowmik, S.K.Parui and U.Roy [13] proposed skew detection algorithm which utilizes the candidate path for a measurement of skewness. The algorithm relies on the fact that the angle between two successive points of the candidate path would give the amount of skew between the two points. Mathematical morphological and weighted least squares method is used to correct handwritten baseline skew in [33]. In [35] a novel skew detection algorithm and the robustness skew correction algorithm is

proposed. In [36] to correct skewed word, horizontal projection of the word and the Wigner-Ville Distribution of projection are done. The idea is based upon the fact that the envelop of the histogram is getting smoother when the word tends to be oriented at the vertical position i.e. $\pm$ 90 degree.

For slant correction Slant normalization algorithms are designed to straight the strokes up in [27].The slant normalization algorithm is modified form of algorithm proposed by Kim and Govindraju.   It takes a chain code representation of the character strokes. The slant angle is calculated from tracing the strokes aligned within an angle around vertical direction. In [28] proposed a method for slant estimation correction by using chain code method. Seiichi Uchida, Eiji Taira, and Hiroaki Sakoe proposed a nonuniform slant correction technique in [29] where the slant correction problem is formulated as an optimal estimation problem of local slant angles at all horizontal positions. The optimal estimation is governed by a criterion function and several constraints designed to evaluate global and local goodness of the estimated local angles. The optimal local slant angles which maximize the criterion function satisfying the constraints are searched for efficiently by a dynamic programming (DP)-based algorithm.   In [30] R.Bertolami, S.Uchida, M.Zimmermann, H. Bunke proposed a non-uniform slant correction is method assuming an $M_{(width)} \times N_{(height)}$ text line image. In [31] Jian-xiong Dong and Dominique Ponson Adam Krzy˙zak and Ching Y. Suen proposed a method for skew and slant correction based on Radon transform. The methods of radon transform which is provides better result for the skew correction.

B.   Feature extraction

Feature extractor is a vital part of any recognition system. The main intend of feature extraction is to depict the pattern by means of bare minimum number of attributes. One significant job in design of pattern recognition system is to develop as algorithm to extort characteristics of pattern from initial measurement. Some features that have been carried for numeral recognition are geometric feature, topological, directional, mathematical & structural features. Feature selection is a process of minimizing the number of features and maximizing the discriminating property of the feature set.   Feature selection is a process that aims to identify an optimal subset of relevant features from a large number of features collected in the data set, such that the overall accuracy of classification is increased. Study of number of papers concern with the handwritten script recognition; indicate that statistical and structural features are mostly used

feature [47]. With these features number of researchers also proposed some of the different features. The features should be selected in such a way that it reduces the intra-class variability and increases the inter-class discriminability in the feature space [46]. In [23] some features consider which are Dots, Junction, Endpoints, Loops and to extract those features from the image a sliding window is applied, which creates a sequence of related feature vectors. To exploit the sequential nature, a model based on HMM is proposed.

Sriganesh Madhvanath in [17] proposed structural features which are best suited for holistic recognition of handwriting. Higher level structural features, includes edges and end points, and perceptual features, and includes dots, holes, ascenders, descenders, and t-bars. In [14] scalar features which includes the height , width , aspect ratio , area , number of descenders in the word , number of ascenders in the word as well as   profile based features such as projection profile, upper word profile, lower word profile   are proposed.

In [24], a foreground pixel distribution feature is considered where the feature extraction algorithm subdivides the character image recursively, Zone Density feature where the foreground pixels are divide in to zones and zoning density is find. Another Background Directional Distribution feature is consider where directional distribution of neighboring background pixels to foreground pixels, and directional distribution values are calculated.   . In [9] Ramteke used the features which are based on invariants moment technique which is used to evaluate seven distributed parameter of a character image. The moment invariants are known to be invariant under translation, scaling, rotation and reflection. They are measures of the pixel distribution around the centre of gravity of the character and allow capturing the global character shape information.   Ramteke evaluated the moment invariants using central moments of the image function $f(x, y)$ up to third order. Central invariant moments which are used as features.   The invariant moments are evaluated by taking log of the absolute value of the moment of each of the sample image of the numerals. Also mean and standard deviation are determined for each feature. For feature extraction techniques two different types of features i.e. sum graph and HMM features are consider by Mohit Mehta, Rupesh Sanchati and Ajay Marchya in Automatic Cheque Processing System. Shirorekha detection, vertical spine detection, number of intersections, accumulated directional gradient change feature has been considered as the features in [15] by S. Arora and others. In [48] S. Arora and others Combine Multiple Feature Extraction Techniques for Handwritten

three feature sets are used. Which includes shadow features which extracted from scaled bitmapped character image where rectangular boundary enclosing the character image is divided into eight octants and for each octant shadow of character segment is computed on two perpendicular sides to obtain shadow features. Shadow feature is basically the length of the projection on the sides and this feature computed on scaled image. Eight different features are consider in [44] which are Bottom-max-row-no, Top-horizontal-line Tick-component, Bottom-component, Top-pipe-size, Bottom-pipe-size, Top-pipe-density, Bottom-pipe-density S. Kumar [45] has compared performances of feature-extraction methods on handwritten characters.

The various features covered are Kirsch directional edges, distance transform, chain code, gradient and directional distance distribution. Among all methods it is found that Kirsch directional edges are least performing and gradient is best performing with SVM classifiers. With multilayer perceptrons (MLP), the performance of gradient and directional distance distribution is almost same. The chain-code-based feature is better as compared to Kirsch directional edges and distance transform. A new feature is also proposed by Kumar where the gradient direction is quantized into four-directional levels and each gradient map is divided into 4×4 regions. In [51] a texture feature extraction approach is followed where a multiresolution approach based on a discrete wavelet transform (DWT) for the texture feature extraction. Three types of features are used in Devanagari Numeral recognition in [52] proposed by R.Bajaj and others which are density features, moment feature of right, left, upper and lower profile curves, and descriptive component features. S. Shelke, S. Apte proposed extracted features such as pixel density feature, Euclidean distance feature and modified approximation wavelet features in [49].

Zernike moments are invariant to global features for images and hence Zernike moments have been analyzed and implemented. Zernike Moments are used a feature vector for recognizing hand printed script in [50]. Various authors have proposed number of feature and feature extraction techniques till the date in various journal papers and conferences word wide.

C. Recognition / Classification

Recognition or classification of a particular data is depend on the selection of the feature and a classifier which can classify or recognize a particular data pattern of that class. Classifier is called 'heart' of the pattern recognition system. It takes feature vector and assigns to it a decision where decision is the number of pattern class which the pattern recognizer has decided. Various classifiers are proposed in the basic application of pattern recognition.

Different classifiers are tested in the [53] which includes K-Nearest –Neighbor (KNN) classifier , which is an effective technique for classification problem in which the pattern classes exhibits a reasonably small degree of variability. The K-NN classifier is based on the assumption that the classification of an instance is most similar to the classification of other instances that are nearby in the vector space. It works by calculating the distances between one input patterns with the training patterns. A K-NN classifier takes into account only the K nearest prototypes to the input pattern, and the majority of class values of the K neighbors determine the decision. Another classifier is the Quadratic discriminant function (QDF) of an n-dimensional feature vector. This classifier becomes optimal in the Bayesian sense for normal distributions with known parameters. In this case the QDF achieves the minimum mean error probability, monotonicity i.e. the average error rate of the QDF decreases monotonically with an increase of the feature size. Also a Support Vector Machine is supervised Machine Learning technique. It is primarily a two class classifier. Width of the margin between the classes is the optimization criterion, i.e. the empty area around the decision boundary defined by the distance to the nearest training pattern. These patterns called support vectors, finally define the classification function [53]. In [54] have conducted experiments with the handwritten characters using multilayer perceptron (MLP) and Support Vector Machine (SVM) classifier.

and Apte [49] use the artificial neural network approach. They have used multilayer perceptron (MLP). The MLP is a fully connected network, with an input layer, hidden layer and an output layer, where every neuron in a layer is connected to each and every neuron in the next layer by a weighted link through which the state of the neuron is transmitted. In [25] also use the SVM because of its many attractive features and promising empirical performance especially in classification and nonlinear function estimation. The K- Nearest Neighbor (K-NN) classifies an unknown sample based on the known classification of its neighbors is also used. The curvelet with K-NN gives better results than SVM. In [55] a HMM classifier is proposed where HMM parameters estimates are fine tuned. In [51] K-NN classifier is used for script identification. Numbers of experiments observations are obtained on different languages script. Minimum Edit Distance Method is used in [56] where regular matching results in accuracy of 70-75%.

In [50] support vector machine are used. Their common

factor is the use of technique known as the 'kernel trick' to apply linear classification techniques to non linear classification problems. Here experimental results are obtained for SVM using Zernike Features. N. Sharma, U. Pal, F. Kimura, and S. Pal used Modified Quadratic Discriminant Function (MQDF) in [57]. In [24] used Support Vector Machine classifier which is a supervised Machine Learning Technique. M. Mehta, R. Sanchati and A. Marchya adopt character recognition system based on projection histogram technique which is fed to a neural network. The probabilistic neural network gives the statistical mean and standard deviation of each value. The nearest mean classifier is used as knowledge based classifier. [58] In this paper, the classification accuracy of different classifiers is discussed. The authors used the Euclidian Distance (ED) between the input pattern and mean vector; Projection Distance (PD) which gives the distance from the input pattern to the minimum mean square error hyper plane that approximates the distribution of the sample; Subspace Method (SM) ; Modified Projection Distance (MPD); Parzen Window Classifier (PWC); A modified Quadratic Discriminant function (MQDF), GLVQ Algorithm(GLVQ); Polynomial Classifier(PC); Fisher Linear Discriminant(FLD); Linear and quadratic Classifiers(LC,QC); Nearest Neighbor Rule Based Classification Methods Nearest Neighbor Classification (NNC); Support Vector Machines ; Radial Basis Function Networks; Hidden Markov Model (HMM).

R.Bajaj [52] used a neural network-based classification scheme is designed for this task. The numerals are represented by feature vectors of three types. Three different neural classifiers have been used for classification of these numerals. The network architecture includes the following components such as Kohonen self-organizing maps at the lowest layer; A single-layer super-structure laid on Kohonen nets; Multi-layer-perception (MLP)-based classifier for the segment features; Meta-pi combining net for integration. Banashree and Vasanta attempted to make use of Neuro memetic model to recognize the handwritten number in pattern recognition area. In this mean absolute percentage error is minimized by adjusting the inpur/output parameter which is the main aim for the training neutal network. Mematic algorithm with operator like arithmetic crossover and non uniform mutation are used to refrain the parameters of neural network and finally evaluation is done by using fitness function. A two class classification approach is used in [59] where the Matlab toolbox for pattern recognition is used for training two class classifiers for zone classification.

IV. OBSERVATION

The classification time of SVM classifier is very high as compared to MLP but the recognition error is low It is essential to divide the Devanagari dataset into subsets so that the classification time may be reduced without compromising accuracy. The characters of each subset, so obtained, are further recognized using SVM classifier. The recognition rate on Devanagari is 94.2% using DFS-345 features in combination with the three stage recognition scheme having SVM as classifier for all the three stages. The recognition error with MLP classifier in any stage is large as compared to SVM. Quadratic Classifier gives accuracy 94.65%. KNN-(Correlation and K=3) gives accuracy 96.65%, SVM (γ=1.1 and c=500) gives accuracy 98.49 % [54]. Curvelet feature gives accuracy of 93.21 and error rate 6.79 for K-NN and for SVM accuracy of 85.6 and error rate of 14.4 [25] chain-code histogram-based features, when confidence score of the top choice recognition label is as high as 95%, 90%, 80%, 70%, 60%, 50% respectively [61].

Three MLP's are designed for features namely Chain code Histogram based, four side views based and Shadow based features. Results of three MLP's are combined using weighted majority scheme discussed above. Combined MLP is giving 98.61% accuracy [48]. SVM gives test set with 92.38% and training set 99.62% Multiple Neural Network Classifier Combination (ANNs) gives the Min test set of 78.49% and training set of 91.08% and Max test set of 93.93% and training set of 98.24% Weighted gives 90.44%test set and 97.94% training set majority gives 99.08% and training set of 97.94%. Performance of different K- NN gives recognition rate as 85.8, 85.3, 85.3 and error rate 14.2, 11.5, 8.5 with reject rate of 0, 3.7, 8.2 [62].The proposed scheme for numerals is 98.86% and for characters is 80.36% with zero percent rejection. 99.80% accuracy was obtained if we consider first two top choices of the recognition results for numerals and 90.56% accuracy was obtained considering the first two top choices for characters [57].

In the recent past, Department of Information Technology (DIT), Government of India formed a Consortium of several Institutions/Universities of India involved in OCR activities and provided considerable amount of fund to this Consortium to improve the quality of research related to Indian language OCR [8]. Skew detection will be difficult in the absence of shirorekha as some of the existing works related to skew detection are based on the presence of shirorekha on words. The straightness of the shirorekha is also an issue of concern [8]. R M K Sinha of IIT Kanpur has been instrumental in the development of Indian language recognition and processing since the beginning [12].

V. CONCLUSSION

Only a few papers are published on script identification. Generally researchers assume that a given document is written in a specific script. In countries like India, where many languages and scripts exist, the identification of script has to be done prior to the recognition in applications like postal address reader, where address can be written in any Indian script. More research toward this direction on handwritten documents is expected in near future. 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, but not much work has been done towards off-line handwriting recognition of Devanagari script. To get the idea of the recognition results of different classifiers and to provide new benchmark for future research, Some of the leading institutes in India doing research in Devanagari OCR are Indian Statistical Institute at Kolkata, International Institute of Information Technology at Hyderabad, Indian Institute of Science at Bangalore, and Indian Institute of Technology at New Delhi.

In India huge volumes of historical documents and books (handwritten or printed in Devanagari script) remain to be digitized for better access, sharing, indexing, etc. This will definitely be helpful for other research communities in India in the areas of social sciences, economics, and linguistics. From the survey, it is noted that the errors in recognizing printed Devanagari characters are mainly due to incorrect character segmentation of touching or broken characters. Because of upper and lower modifiers of Devanagari text, many portions of two consecutive lines may also overlap and proper segmentation of such overlapped portions are needed to get higher accuracy. Many authors suggest that the post processing of classifier outputs by integrating a dictionary with the OCR system can significantly reduce the misclassifications in printed as well as handwritten word recognitions.

REFERENCES

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

[2] U. Pal, R. K. Roy, K. Roy, and 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] K.-Roy, A. Banerji and U.Pal, "A System for Word-wise Handwritten Script Identification for Indian Postal Automation", IEEE INDIA annual conference 2004, INDLCON 2004.

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

[5] S.Marinai, and H.Fujisawa, "Machine Learning in Document Analysis and Recognition", In Springer-Verlag, Berlin Heidelberg, 2008.

[6] M.Mehta, R.Sanchati and A. Marchya, "Automatic Cheque Processing System", International Journal of Computer and Electrical Engineering, Vol. 2, No. 4, August 2010.

[7] N.Otsu, "A threshold selection method from gray-level histograms" IEEE transactions on systems, man, and cybernetics, vol. smc-9, no.1, January 1979.

[8] R.Jayadevan, S.R.Kolhe, P.M.Patil, U.Pal, "Automatic processing of handwritten bank cheque images: a survey", Springer IJDAR July 16, 2011.

[9] R.J.Ramteke, "Invariant Moments Based Feature Extraction for Handwritten Devanagari Vowels Recognition" 2010 International Journal of Computer Applications (0975 - 8887) Volume 1 – No. 18.

[10] R. Jayadevan, S. R. Kolhe, P.M. Patil, and U.Pal, "Offline Recognition of Devanagari Script: A Survey" IEEE transactions on systems, man, and cybernetics—part c: applications and reviews. Nov 17, 2010.

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

[12] R. Mahesh, K. Sinha Indian Institute of Technology, Kanpur, "A Journey from Indian Scripts Processing to Indian Language Processing" IEEE Annals of the History of Computing January–March 2009.

[13] A.Roy, T.K.Bhowmik, S.K.Parui and U.Roy, "A Novel Approach to Skew Detection and Character Segmentation for Handwritten Bangla Words" Proceedings of the Digital Imaging Computing: Techniques and Applications. DICTA 2005.

[14] Victor Lavrenko, T.M. Rath, R. Manmatha, "Holistic Word Recognition for Handwritten Historical Documents". In Proceedings of First intentional workshop on Document image analysis for libraries, august 2004: page 278-287.

[15] S.Arora, L.MaliK, D.Bhattacharjee, M.Nasipuri, "A novel approach for handwritten Devanagari character recognition". ICSIP 30 June 2010.

[16] M.M.Haji, "Farsi handwritten word recognition using continuous hidden markov models and structural features". Jan 2005, MSC thesis, Shiraj University, Iran.

[17] S.Madhvanath, V.Govindaraju, "The role of holistic paradigms in handwritten word recognition" IEEE transactions on pattern analysis and machine intelligence, vol. 23, no. 2, February 2001.

[18] U.Pal, B.B.Choudhuri, "Script Line Separation from Indian Multi-Script Documents," 5th Int. Conference on Document Analysis and Recognition (IEEE Comput. Soc.Press) 1999, pp.406-409

[19] U. Pal, S. Sinha and B. B. Chaudhuri : Multi-Script Line identification from Indian Documents, In Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR 2003) 2003 IEEE, vol.2, 2003 pp.880-884,

[20] S. Chaudhury, R. Sheth, "Trainable script identification strategies for Indian languages", In Proc.5th Int. Conf. on Document Analysis and Recognition (IEEE Comput. Soc. Press), pp. 657–660, 1999.

[21] R. M. K. Sinha, "A Syntactic pattern analysis system and its application to Devanagari script recognition," Ph.D. Thesis, Dept. Elect. Eng., Indian Institute of Technology, Kanpur, India, 1973.

[22] B.Shaw, S.Parui, M.Shridhar, "Offline Handwritten Devanagari Word Recognition: A Segmentation Based Approach". IEEE international Conference on information technology, Dec 2008: page 256-257.

[23] D.Modolo,T.Kooi, "Off-line cursive handwritten word recognition using hidden Markov Models". Available at *www.ai-project-11.googlecode.com/svn.*

[24] M.Jangid, "Devanagari Isolated Character Recognition by using Statistical features" International Journal on Computer Science and Engineering (IJCSE) ISSN: June 2011, Vol. 3, 6 June 2011.

[25] B.Singh, A.Mittal, M.A.Ansari, "Handwritten Devanagari Word Recognition: A Curvelet Transform Based Approach" International Journal on Computer Science and Engineering (IJCSE). ISSN: Vol. 3. Apr 2011

[26] H.Swethalakshmi, A.Jayaraman, V.S.Chakravarthy, C.C.Sekhar, "Online Handwritten Character Recognition of Devanagari and Telugu Characters using Support Vector Machines". Oct 2006.

[27] X. Liu and Z. Shi, "A Format-Driven Handwritten Word Recognition System", In Seventh international conference on Document Analysis and Recognition, 6 Aug 2003.

[28] Y.Dingt, F. Kimurat, Y.Miyaket, M.Shridhartt, "Accuracy Improvement of Slant Estimation for Handwritten Words". In proceedings of 15th intentional conference on pattern recognition, 2000: vol. 4 pp. 527-530.

[29] S.Uchida, E.Taira, H.Sakoe, "Nonuniform Slant Correction Using Dynamic Programming". In proceedings of sixth intentional conference on Document analysis and recognition, 2001:pp.434-438.

[30] R. Bertolami, S. Uchida, M.Zimmermann, Horst Bunke, "Non-Uniform Slant Correction for Handwritten Text Line Recognition", Ninth IEEE International Conference on Document Analysis and Recognition (ICDAR) 2007: pp.18-22.

[31] J.Dong, D.Ponson, A. Krzyzak, C.Y.Suen, "Cursive word skew/slant corrections based on Radon transform." In Eighth International conference on Document Analysis and Recognition, 29Aug 2005: page 478-483.

[32] R.kapoor, D. bagai, T.S. kamal, "Skew angle detection of a cursive handwritten Devanagari script character image" J. Indian Inst. Sci., May-Aug. 2002, Vol: 82, pp.161–175 © Indian Institute of Science.

[33] M.E.Morita, J.Facon, F.Bortolozzi, "Mathematical Morphology and Weighted Least Squares to Correct Handwriting Baseline Skew" National Scientific Research Council, number 520324/96. Proceedings of the fifth International Conference on Document Analysis and Recognition, Sep1999: page 430-433

[34] S. Madhvanath' and V. Govindaraju, "Contour-based Image Preprocessing for Holistic Handwritten Word Recognition", In Proceedings of fourth intentional conference on Document analysis and recognition, 1997: Vol 2, pp. 536-539.

[35] M.L.M Karunanayaka, C.A Marasinghe, N.D Kodikara, "Thresholding, Noise Reduction and Skew correction of Sinhala Handwritten Words" MVA2005 IAPR Conference on Machine VIsion Applications, May 16-18, 2005 Tsukuba Science City, Japan.

[36] Kavalliera, U.N.Fakotakis, G.K.Akis, "New algorithms for skewing correction and slant removal on word-level**,** IEEE Proceedings of 6th intentional conference on Electronics, Circuits and Systems, Sep 1999: Vol 2, pp. 1159-1162.

[37] M.Blumenstein, C.K.Cheng, X. Y. Liu, "New preprocessing techniques for handwritten word recognition". Proceedings of Visualization, Imaging, and Image processing 2002.

[38] A.Rehman, D.Mohammad, G.Sulong, T.Saba, "Simple and effective techniques for core-region detection and slant correction in offline script recognition"2009 IEEE International Conference on Signal and Image Processing Applications.

[39] P.Slavik, V.Govindaraju, "Equivalence of Different Methods for Slant and Skew Corrections in Word Recognition Applications" IEEE transactions on pattern analysis and machine intelligence, vol. 23, no. 3, march 2001.

[40] S. Airphaiboon, M. Sangworasil, S.Kondo, "Off-line Handwritten Thai Characters from Word Script." In proceedings of the 12th IAPR international conference on pattern recognition, Oct 1994, Vol 2, pp.445-449.

[41] G.Agrawal, Kshitij, A.Mukerjee, N.Kumar, "Handwritten Devanagari Script Segmentation using Support Vector Machines". *www.geocities.ws/kumarnimit/pubs/IJCNN04.pdf*

[42] V.Lavrenko, T.M. Rath and R. Manmatha, "holistic word recognition for handwritten historical documents" Center for Intelligent Information Retrieval and in part by the National Science Foundation under grant number IIS-9909073 and in part by SPAWARSYSCEN-SD grant number N66001-02-1-8903. Proceedings First international workshop on Document Image Analysis for Libraries, 2004: pp 278-287.

[43] N. P. Banashree, and R. Vasanta, "OCR for Script Identification of Hindi (Devanagari) merals using Feature Sub Selection by Means of End-Point with Neuro-Memetic Model" World Academy of Science, Engineering and Technology 2007.

[44] M.C.Padma, P.A.Vijaya, "Identification of Telugu, Devanagari and English scripts using discriminating features" International Journal of Computer science & Information Technology (IJCSIT), Vol 1, No 2, November 2009.

[45] S. Kumar, "Performance comparison of features on Devanagari handprinted dataset," Int. J. Recent Trends, vol. 1, no. 2, pp. 33–37, 2009.

[46] B.Shaw and S. Parui, "A Two Stage Recognition Scheme for Offline Handwritten Devanagari Words" World Scientific Review Volume - February 23, 2008.

[47] S.Madhvanath, V.Govindaraju, "The Role of Holistic Paradigms in Handwritten Word Recognition" IEEE Ttransactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 2, Feb.2001.

[48] S. Arora, D.Bhattacharjee, M.Nasipuri, D.K.Basu, M.Kundu, "Combining Multiple Feature Extraction Techniques for Handwritten Devanagari Character Recognition" 2008 IEEE Region 10 Colloquium and the Third ICIIS, Kharagpur, INDIA December 8-10. 342.

[49] S.Shelke, S. 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.

[50] Thungamani.M, Dr. R.Kumar, P.Keshava, P.Shravani, K.Rau, "Off-line Handwritten Kannada Text Recognition using Support Vector Machine using Zernike Moments" IJCSNS International Journal of Computer Science and Network Security, VOL.11 No.7, July 2011.

[51] P.S.Hiremath, S.Shivashankar, D. Jagdeesh , V Pujari, Mouneswara, "Script identification in a handwritten document image using texture features" 2010 IEEE.

[52] R.Bajaj, L.Dey, S. Chaudhury, "Devanagari numeral recognition by combining decision of multiple connectionist classifiers" Sadhana Vol. 27, Part 1, Feb.2002, pp. 59–72. © Printed in India.

[53] M.Jangid, K.Singh, R.Dhir, R.Rani, "Performance Comparison of Devanagari and written Numerals Recognition" International Journal of Computer Applications (0975 – 8887) Volume 22– No.1, May 2011.

[54] S.Kumar, "A Three Tier Scheme for Devanagari Hand-printed Character Recognition" 2009 World Congress on Nature &

Biologically Inspired Computing (NaBIC 2009).

[55] B.Shaw, S.K.Parui, M.Shridhar, "Offline Handwritten Devanagari Word Recognition: A holistic approach based on directional chain code feature and HMM." International conference on information technology, Dec 2008: pp 203-208.

[56] Dr. P. S. Deshpande , S.Arora , L. Malik, "Fine Classification & Recognition of Hand Written Devanagari Characters with Regular Expressions & Minimum Edit Distance Method" JOURNAL OF COMPUTERS, VOL. 3, NO. 5, MAY 2008.

[57] N.Sharma, U. Pal, F. Kimura and S. Pal, "Recognition of Off-Line Handwritten Devanagari Characters Using Quadratic Classifier", Springer-Verlag Berlin Heidelberg 2006: ICVGIP pp. 805-816.

[58] A.N.Holambe, Dr.R.C.Thool, "Comparative study of different classifiers for Devanagari handwritten character recognition" International Journal of Engineering Science and Technology Vol. 2 (7), 2010:pp.2681-2689.

[59] J.Kumar, "codebook based handwritten and printed Arabic text zone classification", Available at: http://*www.cs.umd.edu*.

[60] S.Arora. D.Bhattacharjee, M.Nasipuri, L.Malik, M.Kundu, D.K.Basu, "Performance Comparison of SVM and ANN for Handwritten Devanagari Character Recognition. IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 3, No 6, May 2010.

[61] S.Chand, S.Pal, K.Franke, U.Pal, "Two-stage Approach for word-wise Script Identification" 10th International Conference on Document Analysis and Recognition, 2009.

[62] De Stefano, A. D.Cioppa, A. Marcelli, "Handwritten Numeral Recognition by means of Evolutionary Algorithms". Proceedings of the fifth international conference on document analysis and recognition Sept 1999. Pp. 804-80.

[63] Gonzalez R.C., Woods R.E., "Digital Image Processing",Prentice hall, 2002.

[64] Duda R.O., Heart P.E., Stork D.G. , "Pattern Classification", Wiley, 2001.

Ashwin Ramteke received the B.E. Degree in Electronics and Telecomm from Amravati University, Amravati in 2007. Worked as a lecturer at SGSITS, Indore from 2007 to 2010. A Student Member of IEEE in 2012. Currently a final year student of M.E. in Electronics and telecom (signal processing) from VIT, Pune.

Milind E Rane received his BE degree in Electronics engineering from University of Pune and M Tech in Digital Electronics from Visvesvaraya Technological University, Belgaum, in 1999 and 2001 respectively. His research interest includes image processing, pattern recognition and Biometrics Recognition.