Survey Paper on Handwritten Devanagari Numeral Recognition

Madhav Goyal[1], Naresh Kumar Garg[2]

[1]Department of Computer Science & Engg, Guru Ram Dass Institute of Engg. & Technology, Bathinda
[2]Department of Computer Science & Engg, GZSPTU Campus, Bathinda
[1] Madhav.jaito@gmail.com, [2]Naresh2834@rediffmail.com

Abstract-**In this paper we are going to present a survey of the various research papers published by corresponding authors in the field of OCR of Devanagari script by using different feature extraction techniques. This paper will provide complete guidance and update to the readers, working in the Devanagari Optical Character Recognition area.**

Keywords- Devanagari Character Recognition, Off-line Handwriting Recognition, Online Handwriting Recognition.

1. INTRODUCTION

Handwritten recognition is a challenging and difficult task in pattern recognition due to its complex structural properties. The main reason of these complex structures is that the different writers have different writing styles. In case of printed Hindi numerals near about 100% accuracy is obtained but in case of handwritten numerals/characters the main obstacle is the variations in the writing styles of different writers. Hindi is the national language of India which is based on Devanagari script Therefore; handwritten OCR of Devanagari script is an area of active research in order to increase the functionality of OCR system. Numeral recognition systems have various applications aautomatic reading of numerical fields has been introduced in various fields such as:

1. Processing bank check amounts.
2. Numeric entries in forms filled up by hand.
3. Recognize zip codes on mail for postal address sorting.

This Paper is structured as Follows: Section II reprsents the Database of Hindi numerals written by different writers and will also illustrate the variation in the writing styles of different writers and Section III describes the literature survey on various Feature Extraction techniques.

1.1 Types of Handwriting recognition

Handwriting recognition is classified as Online Handwriting Recognition and Offline Handwriting Recognition which are discussed as under:

1. Online Handwriting Recognition: In online recognition systems, the system recognizes continuously as the writer writes the data. It basically goes along the writing process.
2. Offline Handwriting Recognition: Offline handwriting recognition is performed after the writing or printing is complete.

1.2 Problems in Handwritten Recognition:

Sometimes the Numerals are classified wrongly due to many reasons which are described as under:

1. Numeral is not properly written.
2. Some Numerals have similar structures.
3. Sometimes two or more numerals may overlap.
4. Sometimes system confuses the numeral with some other numeral and does not recognize it correctly.

2. DATABASE

The database is constructed by taking handwritten data from 25. Each writer was asked to write digits from 0 to 9 in Devanagari Script. The Writers for writing a data were taken from various backgrounds like school students, College students, Teachers, Shopkeepers, Housewives, engineers and many more. Total digit images are 250 in number. Figures 1a & 1b contains some part of handwritten digits database.
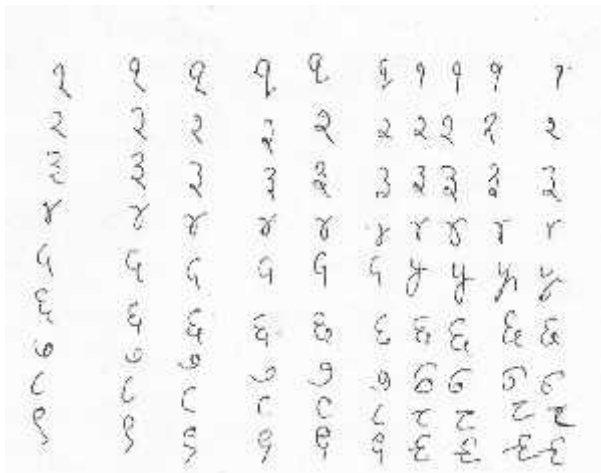

Figure 1a: Part of database

Figure 1b: Part of database

There is large number of variations in the shapes of the numerals written by different writers which is shown in figure 2.
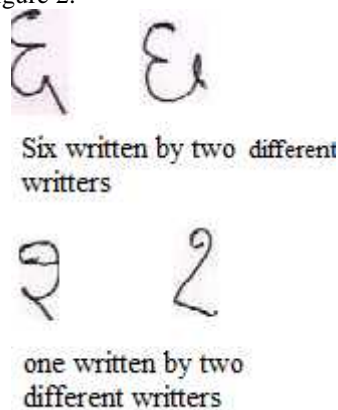

Figure 2: Variations b/w the writing of different writers

3. LITERATURE SURVEY

The work on handwritten Devanagari numeral is carried by **M. Hanmandlu, A.V. Nath, A.C. Mishra and V.K. Madasu [1]** proposed a recognition which is based on the modified exponential membership function fitted to the fuzzy sets derived from features consisting of normalized distances obtained using the Box approach. The exponential membership function is modified by two structural parameters that are estimated by optimizing the entropy subject to the attainment of membership function to unity. The optimization strategy used is the foraging model of E.coli bacteria. Experimentation is carried out on a limited database of nearly 3500 samples. The overall recognition is found to be 96%. **Alessandro L. Koerich, Pedro R. Kalva [2]** have developed three classification strategies to recognize unconstrained handwritten characters to accomplish this task they used , four multilayer perceptron classifiers (MLP) were built and used into three different classification strategies: combination of two 26-class classifiers; 26-metaclass classifier; 52-class classifier. The recognition rate achieved by the metaclass classifier

(87.8%) outperforms the other approaches (82.9% and 86.3%). **Sethi and Chatterjee[3]** have done some studies on Devnagari script and presented a Devnagari hand-printed numeral recognition system based on binary decision tree classifier. **Shailedra Kumar Shrivastava, Sanjay S. Gharde [4]** used Moment Invariant and Affine Moment Invariant techniques as feature extractor. 18 features are extracted from each image by using these techniques which is used in Support Vector Machine for recognition purpose. Binary classification techniques of Support Vector Machine is implemented and linear kernel function is used in SVM. This linear SVM produces 99.48% overall recognition rate which is the highest among all techniques applied on handwritten Devanagari numeral recognition system. **Prerna Singh and Nidhi tyagi[5]** proposed a method for recognition of Handwritten Devanagari numerals using Radial basis function feed forward neural network that computes activation at the hidden neuron in a way that is different from character & numeral recognition system consist of three main components, they are pre-processing, feature extraction and classification. **Sandhya Arora, Debotosh Bhattacharjee, Mita Nasipuri and Latesh Malik[6]** presented a two stage classification stratergy for handwritten characters written in Devanagari script. In first stage structural properties like shirorekha, spine in character are used and In second stage some intersection features of characters which are fed to a feedforward neural network are exploited. Histogram based method does not work for finding shirorekha, vertical bar in handwritten devnagari characters. So they designed a differential distance based technique to find a near straight line for shirorekha and spine. This approach has been tested for 50000 samples and got 89.12% success. **J.Pradeep , E.Srinivasan and S.Himavathi [7]** proposed handwritten character recognition system using multilayer Feed forward neural network in which Three different orientations, namely, horizontal, vertical and diagonal directions are used for extracting 54 features from each character. The diagonal orientation for feature extraction is identified to be the most suitable method as it yields higher recognition accuracy of 98% for 54 features and 99% for 69 features. **Brijmohan Singh , Ankush Mittal and Debashis Ghosh[8]** used two different methods for extracting features from handwritten Devnagari characters, the Curvelet Transform and the Character Geometry, and compare their recognition performances using two different classifiers which are Support Vector Machine (SVM) with Radial Basis Function (RBF), and the k-Nearest Neighbour (k-NN) classifier. Different classification accuracy measures, such as True Positive (TP) Rate, False positive (FP) Rate, Precision, Recall and F- Measure, are used for the purpose. Results obtained show that Curvelet features with k-NN classifier performs the best, yielding accuracy as high as 93.8%. **A. Elnagar, F. Al-Kharousi, and S. Harous [9]** proposed a method for the recognition of handwritten Hindi numerals which is based on structural descriptors of the shapes

of numerals. In this method first of all, hand written numeral is scanned, normalized and thinned after this pre-processing stage the pre-processed image is passed through the algorithm which segments the scanned image into strokes to extract features and results the syntactic representation of numeral image and then recognition is performed based on all structural descriptors obtained above and resulted the recognition rate 90 to 94%.**Anilkumar N. Holambe, Dr.Ravinder.C.Thool and Dr.S.M.Jagade[10]** have extracted the features of handwritten and printed characters and numerals of Devanagari script. They extracted the Gradient features of the Devanagari script by using two operators i.e. Sobel and Robert operator respectively .They computed gradient feature in 8,12,16,32 directions and got different feature vectors respectively. By using 8-directional Sobel operator 94% accuracy was obtained for handwritten Characters, 94.2% was obtained for handwritten numerals, 98% accuracy was obtained for Printed Characters and 97% was obtained for Printed numerals , By using 12-directional Sobel operator accuracy 94.76% was obtained for handwritten Characters, 94.44% was obtained for handwritten numerals, 98% accuracy was obtained for Printed Characters and 97% was obtained for Printed numerals, By using 16-directional Sobel operator accuracy 96% was obtained for handwritten Characters, 95% was obtained for handwritten numerals, 98.45% accuracy was obtained for Printed Characters and 98.05% was obtained for Printed numerals and By using 32-directional Sobel operator accuracy 97% was obtained for handwritten Characters, 96% was obtained for handwritten numerals, 98.78% accuracy was obtained for Printed Characters and 98.0% was obtained for Printed numerals. By using 8-directional Robert operator 94.45% accuracy was obtained for handwritten Characters, 94% was obtained for handwritten numerals, 97% accuracy was obtained for Printed Characters and 96% was obtained for Printed numerals , By using 12-directional Robert operator accuracy 95.06% was obtained for handwritten Characters, 96% was obtained for handwritten numerals, 98% accuracy was obtained for Printed Characters and 95% was obtained for Printed numerals, By using 16-directional Robert operator accuracy 95.67% was obtained for handwritten Characters, 95% was obtained for handwritten numerals, 98.02% accuracy was obtained for Printed Characters and 96.05% was obtained for Printed numerals and By using 32-directional Robert operator accuracy 96.09% was obtained for handwritten Characters, 97% was obtained for handwritten numerals, 98.0% accuracy was obtained for Printed Characters and 97.0% was obtained for Printed numerals. Here the accuracy obtained by Sobel operator is high as compared to Robert operator. **Banashree N. P., and R. Vasanta[11]** proposed a recognition scheme for handwritten Hindi (devnagiri) numerals by using global based feature extraction approach using end-points information, which is

extracted from images of isolated numerals. These feature vectors are fed to neuromemetic model that has been trained to recognize a Hindi numeral. In proposed scheme data sets are fed to neuromemetic algorithm, which identifies the rule with highest fitness value of nearly 100 % & template associates with this rule is nothing but identified numerals. Therecognition rate by using this approach is 92-97 %.**U. Pal, A. Belaid and B. B. Chaudhuri[12]** proposed the scheme which is based on the features obtained from the concept of water overflow from the reservoir as well as topological and structural features of the numerals. The proposed scheme is tested on data collected from different individuals of various background and obtained an overall recognition accuracy of about 92.8% from12000 data. **Rawan I. Zaghloul, Dojanah M.K. Bader Enas and F. AlRawashdeh[13]** proposed an algorithm which was implemented in MATLAB and tested with a large sample of handwritten numeral datasets for different writers in different ages. Pattern recognition techniques are used to identify Hindi (Arabic) handwritten numerals. In this scheme they used various feature extraction techniques like Detection of loops, detection of centroid, horizontal projection and segmentation of numeral image to find useful information. After testing, high accuracy was achieved, it ranges from 95% for some numerals and up to 99% for others. **Akhilesh Pandey, Amresh Kumar, Rajiv Kumar and Amod Tiwari[14]** proposed a majority voting scheme for off-line hand written Hindi numbers recognitions. The main objective of this research is to find out best recognition result using multiple classifiers. The proposed technique uses simple profile and contour base triangular area representation technique for finding feature extraction and majority voting scheme on back propagation and cascade feed forward neural network for classification. The average recognition result of this approach is 94.16%.

Conclusion
The complete review of various feature extraction papers is presented which includes the review of handwritten as well as printed numeral recognition, character recognition and review of comparison of various classifiers is also presented. This paper will act as complete guide to the new readers as well as researchers.

References
[1]   M. Hanmandlu, A.V. Nath, A.C. Mishra and   V.K. Madasu, "Fuzzy Model Based Recognition of Handwritten Hindi Numerals using Bacterial Foraging"

[2]   Alessandro L. Koerich, Pedro R. Kalva, "Unconstrained Handwritten Character Recognition Using Metaclasses of Characters"

[3]   K. Sethi, B. Chatterjee, "Machine recognition of constrained hand-printed Devnagari numerals", J. Inst. Electron. Telecom. Eng. 22 (1976) 532-535

[4]   Shailedra Kumar Shrivastava , Sanjay S. Gharde , "Support Vector Machine for Handwritten Devanagari Numeral Recognition"

[5]   Prerna Singh and Nidhi tyagi, " Radial Basis Function For Handwritten Devanagari Numeral Recognition", *Vol. 2, No. 5, 2011*

[6]   Sandhya Arora, Debotosh Bhattacharjee, Mita Nasipuri and Latesh Malik, "A Two Stage Classification Approach for Handwritten Devanagari Characters"

[7]   J.Pradeep , E.Srinivasan and S.Himavathi, " Diagonal Feature Extraction Based Handwritten Character System Using Neural Network"

[8]   Brijmohan Singh , Ankush Mittal and Debashis Ghosh , "An Evaluation of Different   Feature Extractors and Classifiers for Offline Handwritten Devnagari Character Recognition"

[9]   A. Elnagar, F.Al-Kharousi, and S. Harous "Recognition of Handwritten Hindi Numerals using Structural Descriptors", 1997 IEEE

[10]   Anilkumar N. Holambe, Dr.Ravinder.C.Thool and Dr.S.M.Jagade," Printed and   Handwritten Character & Number Recognition of Devanagari Script using Gradient   Features", International Journal of Computer Applications (0975 – 8887) Volume 2 – No.9, June 2010.

[11]   Banashree N. P., and R. Vasanta, "OCR for Script Identification of Hindi (Devnagari) Numerals using Feature Sub Selection by Means of End-Point with Neuro-Memetic Model", World Academy of Science, Engineering and Technology 4 2007.

[12]   U. Pal, A. Belaid and B. B. Chaudhuri, "A System for Bangla Handwritten Numeral Recognition"

[13]   Rawan I. Zaghloul, Dojanah M.K. Bader Enas and F. AlRawashdeh, "RECOGNITION OF HINDI (ARABIC) HANDWRITTEN NUMERALS", American Journal of Engineering and Applied Sciences, 2012, 5 (2), 132-135 ISSN: 1941-7020

[14]   Akhilesh Pandey, Amresh Kumar, Rajiv Kumar and Amod Tiwari, "Handwritten Devanagri Number Recognition using Majority Voting Scheme", International Journal of Computer Science and Information Technology & Security (IJCSITS), ISSN: 2249-9555 Vol. 2, No.3, June 2012.