

Analysis of Performance Measures for the Development of DEVNAGARI Character Recognition System : A Review

Shrutika C. Wankhade¹, Prashant M. Kakde²

¹ME Student, Department of Electronics and Tele Communication, HVPM's College of Engineering & Technology, Amravati, Maharashtra, India

²Assistant Professor, Department of Electronics and Tele Communication, HVPM's College of Engineering & Technology, Maharashtra, Amravati, India

ABSTRACT

Devanagari script is widely used in the Indian subcontinent in several major languages such as Hindi, Sanskrit, Marathi and Nepali. Recognition of unconstrained (Handwritten) Devanagari character is more complex due to shape of constituent strokes. Hence character recognition has been an active area of research till now and it continues to be a challenging research topic due to its diverse applicable environment. As the size of the vocabulary increases, the complexity of algorithms also increases linearly due to the need for a larger search space. Devanagari script recognition systems using Zernike moments, fuzzy rule and quadratic classifier provide less accuracy and less efficiency. Classification methods based on learning from examples have been widely applied to character recognition have brought forth significant improvements of recognition accuracies. In this work technique like Artificial Neural Network, Support Vector Machines, Particle Swarm Optimization and Genetics Algorithm are implemented for showing the recognized Devnagari character by using Raspberry Pie.

Keywords: Online Character Recognition (OCR), Classifiers, Raspberry Pie, Performance Analysis.

I. INTRODUCTION

Handwritten character recognition is gaining popularity due to its potential application areas which would reduce the task of data entry and save the time in case of form filling, postal automation, and banking etc. Hindi being the National language of India spoken by millions of people and written using Devnagari Script is still in developing stage and lot of work need to be attempted before emerging a satisfactory system with much success rates. But developing a system for handwritten character recognition poses a challenge to the researchers due to the varying shape of the character that may depend upon the writer, the physical and mental condition of the writer, the acquisition device, pen-width, pen ink colour and many other factors. The templates or prototypes in these early methods were designed artificially, selected or averaged from few samples. As the number of samples increased, these simple design methodologies, became insufficient to accommodate the shape variability of samples, and so,

are not able to yield high recognition accuracies. To take full advantage of large volume of sample data, the character recognition community has turned attention to classification methods based on learning from examples strategy, especially based on artificial neural networks (ANNs) from the late 1980s and the 1990s.

New learning methods, using Support Vector Machines (SVMs), Particle Swarm Optimization (PSO) and Genetics Algorithm (GA) are now actively studied and applied in pattern recognition problems. Learning methods have benefited character recognition methods tremendously. They relieve us from painful job of template selection and tuning, and the recognition accuracies get improved significantly because of learning from large sample data. Some excellent results have been reported. Due to the improvements, the problem is far from being solved and efficient accuracies are evolved.

II. LITERATURE SURVEY

An early notable attempt in the area of character recognition research is by Grim Dale in 1959. The origin of a great deal of research work in the early sixties was based on an approach known as analysis-by-synthesis method suggested by Eden in 1968. The great importance of Eden's work was that he formally proved that all handwritten characters are formed by a finite number of schematic features, a point that was implicitly included in previous works. This notion was later used in all methods in syntactic (structural) approaches of character recognition.

Sushama Shelke and P.B.Khanale [4][13] in their paper presents a novel approach for recognition of unconstrained handwritten Marathi characters. The recognition is carried out using multistage feature extraction and classification scheme. The hybrid classifier at the final stage takes the input from two neural network classifiers and template matching classifier and decides the final output based on maximum voting rule. This multistage feature extraction and classification scheme improves the recognition accuracy over individual classifiers considerably. The recognition rate achieved from the proposed method is 95.40%.

Sandhya Arora and Vikas Dongre [5][7] in their paper presents a two stage classification approach for handwritten Devnagari characters the first stage is using structural properties like shirolekha, spine in character and second stage exploits some intersection features of characters which are fed to a feed forward neural network. Simple histogram based method does not work for finding shirolekha, vertical bar (Spine) in handwritten devnagari characters. So they designed a differential distance based technique to find a near straight line for shirolekha and spine. This approach has been tested for 50000 samples and they got 89.12% success.

Yoshimasa Kimura [10] presented a work on how to select features for Character Recognition Using Genetic Algorithm. The author proposes a novel method of feature selection for character recognition using genetic algorithms (GA). Three different classifiers, namely nearest neighbour, k-NN, and SVM were tested

independently to recognize handwritten Devanagari numerals in the performance of SVM in terms of accuracy was better than the other two classifiers. In the feature vector is entered as an input to one of the feed forward back propagation neural network for the classification of handwritten Devanagari characters.

Divakar Yadav, Sonia Sánchez-Cuadrado and Jorge Morato [8] proposed an OCR for printed Hindi text in Devanagari script, by used techniques of Artificial Neural Network (ANN), and then improved recognition efficiency. He performed conversion of gray scaled images to binary images and a back-propagation neural network having two hidden layers is used. The classifier is trained and tested for printed Hindi texts.

R. Bajaj, L. Dey, S. Chaudhari [11], employed three different kinds of features, namely, the density features, Moment features and descriptive component features for classification of Devanagari Numerals. They proposed multi Classifier connectionist architecture for increasing the recognition reliability and they obtained 89.6% accuracy for hand written Devanagari numerals.

In [18], some problems in designing OCRs for Indian scripts are discussed. The authors presented a data collection tool, a segmentation analysis tool, and a feature selection tool capable of tuning the features used for the classification of a particular font or script of another set. With increasing popularity of digital cameras attached with various handheld electronic devices, some new computational challenges have gained significance recently. One such problem is the extraction of textual information from natural scene images captured by such devices. The extracted text information can be sent to an OCR or to a text-to-speech engine for further processing. Because of the recent surge in digital library projects globally and large scale intensification of digitization efforts, it is expected that almost all of man's knowledge will be available in digital form on the Web in the coming year.

The real-time numeral or character recognition systems are developed by acquiring the images using camera, graphic tablet [19], finger movements [20] etc. In this paper, a real time system for Marathi script is proposed, that reads the characters using the camera. The system is developed using Raspberry Pi and Open CV with Python

programming language. The Raspberry Pi pre-processes the character image, recognizes the characters using template matching and then finally converts the characters into speech. The system developed using Raspberry Pi and Open CV is portable and cheap, moreover it recognizes the characters and converts into speech in real time. The characters used in this system are isolated without any modifiers.

III. PROPOSED WORK

The block diagram of basic character recognition is shown in figure.

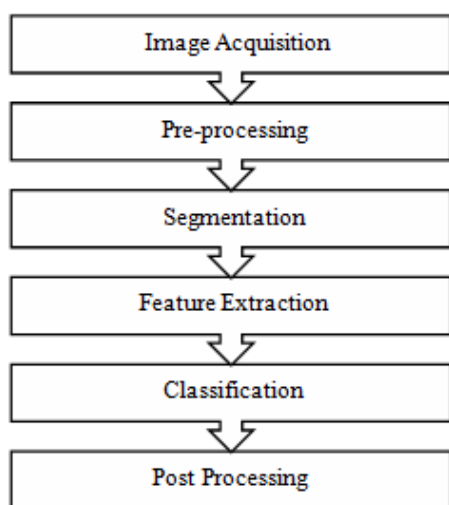


Figure 1: Block Diagram of Character Recognition

A. Pre-Processing

It removes any distortions or discontinuity in the input character and converts the characters in the form recognizable by detection procedure.

B. Classification of Methods

The task of classification is to assign an input pattern represented by feature vectors to one of many pre-specified classes. Here we may use four classifiers described here.

1) *Artificial Neural Network (ANN)*

Artificial neural networks are the computational models that consist of number of simple processing units called neurons distributed in layers namely input, hidden and output that communicate with one another over a large

number of weighted connections. An artificial neural network is based on the operation of biological neural networks. The neurons in the ANN are the electronic counter part of the neurons of the human brain.

2) *Support Vector Machines (SVM)*

Support Vector Machines (SVM's) are a useful technique for data classification. SVM is a supervised learning classifier. A classification task usually involves separating data into training and testing sets. The principle of an SVM is to map the input data onto a higher dimensional feature space nonlinearly related to the input space and determine a separating hyper-plane with maximum margin between the two classes in the feature space. This results in a nonlinear boundary in the input space. The goal of SVM is to produce a model which predicts the target value.

3) *Genetics Algorithm (GA)*

A GAs is an optimization and search method utilized in computer science to find fairly accurate solutions to problems. Gas is generally realized in a computer model, in which a population of runner solutions to an optimization problem progress to better solutions. The evolution starts from a population of completely random. Individuals are occurs in generations. In each generation, the fitness of the entire population is evaluated, and multiple individuals are selected from the present population based on their fitness. These are modified, mutated, or recombined to make a new population, which becomes present in the next iteration of the algorithm. Usually, the solutions are represented in strings of 0s and 1s, though different encodings are also possible. So, evolutionary algorithms play on populations, in its place of coming to one solution.

4) *Particle Swarm Optimization (PSO)*

PSO is a population-based searching method. PSO is a met heuristic technique, as it makes a few or no assumptions about the problem but will search very large spaces of the solution of parameter selection can have a large impact on optimization performance. They can also be tuned by using another overlaying optimizer, known as meta-optimization. It contends that the PSO algorithm and its parameters must be chosen so as to properly balance between exploration and exploitation to avoid premature convergence to a local optimum yet still ensure a good rate of convergence to the optimum.

IV. DISCUSSION AND CONCLUSION

The paper discusses in detail all advances in the area of handwritten character recognition. The most accurate solution provided in this area directly or indirectly depends upon the quality as well as the nature of the material to be read. Various techniques have been described in this paper for character recognition in handwriting recognition system. Few research reports are published recently & tried to solve the problem associated with them. In this table 1; there exist Author, Classifiers and their overall Accuracy. A sort comparison is shown between the different methods proposed so far. From the study done so far, it is analysed that the selection of the classification as well as the feature extraction techniques needs to be proper in order to attain good rate in recognizing the character. Studies in the paper reveals that there is still scope of enhancing the algorithms as well as enhancing the rate of recognition of characters by using raspberry pie.

Table 1. Comparisons between Different Techniques

AUTHOR	CLASSIFIERS	RECOGNITION ACCURACY
J. Pradeep, E. Srinivasan [15]	ANN	90%
N. Sharma [22]	SVM	80.36%
Vedgupt Saraf [12]	GA	92.38%
U. Pal [3]	SVM	93.24%
M. A. Abed [21]	PSO	92.39%
Kumar and Singh [16][9]	SVM	80%

V. REFERENCES

- [1] A. B. Author, Anilkumar N. Holambe Dr. Ravinder. C. Thool, —Comparative Study of Devanagari Handwritten and printed Character & Numerals Recognition using Nearest-Neighbor Classifiers, 978-1-4244-5540- 9/102010 IEEE.
- [2] Constantin Anton, CosminStirbu, Romeo, VasileBadea—Automatic hand writer identification using the feed forward neural networks, 978-0-9564263-7/6/2011 IEEE.
- [3] U. Pal, T. Wakabayashi, F. Kimura, “Comparative Study of Devnagari Handwritten Character Recognition using Different Feature and Classifiers”, *10th Intl. Conf. on Document Analysis and Recognition*, pp. 1111-1115, 2009.
- [4] Sushama Shelke Shaila Apte, —A Novel Multi-feature Multi- Classifier Scheme for Unconstrained Handwritten Devanagari Character Recognition, 2010, 12th International Conference on Frontiers in Handwriting Recognition.
- [5] Sandhya Arora Meghnad Saha Debotosh Bhattacharjee Mita Nasipuri Latesh Malik, —A Two Stage Classification Approach for Handwritten Devanagari Characters, International Conference on Computational Intelligence and Multimedia Applications 2007.
- [6] Sreeraj. M Sumam Mary Idicula, —On-Line Handwritten Character Recognition using Kohonen Networks, 2009 World Congress on Nature & Biologically Inspired Computing (NaBIC 2009).
- [7] Vikas J Dongre Vijay H Mankar, —A Review of Research on Devnagari Character, International Journal of Computer Applications (0975 – 8887) Volume 12– No.2, November 2010.
- [8] Sandhya Arora Debotosh Bhattacharjee Mita Nasipuri D. K. Basu —Recognition of Non-Compound Handwritten Devnagari Characters using a Combination of MLP and Minimum Edit Distance, International Journal of Computer Science and Security (IJCSS), Volume (4)
- [9] C. V. Jawahar, Pavan Kumar, S. S. Ravi Kiran, “A Bilingual OCR for Hindi-Telugu Documents and its Applications”, *Seventh Intl. Conf. on Document Analysis and Recognition (ICDAR 2003)*, pp 1-7, 2003.

- [10] Yoshimasa Kimura, "Feature Selection for Character Recognition Using Genetic Algorithm", Fourth International Conference on Innovative Computing, Information and Control 978-0-7695-3873-0/09© 2009 IEEE, 2009.
- [11] Pahuni Saharan and Neha Sahu,"A Review of Handwritten Devanagari Character Recognition System over the Years, Journal of Network Communications and Emerging Technologies (JNCET) Volume 6, Issue 7, July (2016).
- [12] Vedgupt Saraf, D.S. Rao, Devnagari Script Character Recognition Using Genetic Algorithm for Get Better Efficiency Vol. 7, Issue 2, No 1, March 2010.
- [13] P.B.Khanale and S.D. Chitnis —Handwritten Devnagri Character Recognition using Artificial Neural Networkl, Journal of Artificial Intelligence 4(1): 55-62, 2011, ISSN1994-5450.
- [14] Ved Prakash Agnihotri, Offline Handwritten Devanagari Script Recognition, I.J. Information Technology and Computer Science, pp. 37-48, July. 2012.
- [15] J. Pradeep, E. Srinivasan and S. Himavathi, "Diagonal Based Feature Extraction for Handwritten Alphabets Recognition System Using Neural Network", International Journal of Computer Science & Information Technology (IJCSIT), Vol 3, No 1, Feb 2011.
- [16] Mahesh Jangid Kartar Singh, Renu Dhir, Rajneesh Rani, Performance Comparison of Devnagari Handwritten Numerals Recognitionl, International Journal of Computer Applications (0975 –8887) Volume 22– No.1, May 2011.
- [17] Dr. P. S. Deshpande Mrs. Latesh Malik Mrs. Sandhya Arora, Characterizing Hand Written Devanagari Characters using Evolved Regular Expressionsl, 1-4244-0549-1/ 2006 IEEE.
- [18] [18] W. Ahmad, S. A. Hassan, M. A. Irfan, A. Bais, G. M. Hassan, S. A. Ali Shah, M. Khawaja. "Design and implementation of real time LPR system on a fixed point DSP", *6th International Conference on Emerging Technologies*, pp. 159-163, (2010).
- [19] M. M. Farhad. "Real time numeric character recognition system based on finger movements," International Conference on Electrical Information and Communication Technology, pp. 1–5, (2013).
- [20] A. Beltran, S. Mendoza. "Efficient algorithm for real time handwritten character recognition in mobile Devices", *8th International Conference on Electrical Engineering Computing Science and Automatic Control*, pp. 1–6, (2011).
- [21] M. A. Abed (2013)"Simplifying Handwritten Character Recognition using PSO", European academic research, No. 1.
- [22] N. Sharma, U. Pal, F. Kimura, and S. Pal, "Recognition of Off-Line Handwritten Devnagari Characters Using Quadratic Classifier", *ICVGIP 2006, LNCS 4338*, pp. 805 – 816.