ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 1.852

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

Devanagari Handwritten Numeral Recognition Using Probabilistic Neural Network

Abhay S. Lengare[*1], Dr. Suhas S. Patil[2]
[*1,2] Electronics Engineering Department, K.B.Patil college of Engineering Satara, Satara, India
abhaylengare@gmail.com

Abstract

In the last half century, the English character recognition was studied and the results were of such type that's it can produce technology driven applications. But the same approach cannot be used in case of Indian languages due to the nature of complication in terms of structure and computation. "Hindi" the national language of India (written in Devanagri script) is world's third most popular language after Chinese and English. Devanagri handwritten character recognition has got lot of application in different fields like postal address reading, cheques reading electronically. There are several Handwritten numeral recognition have been proposed and evolved during last few decades. But robustness and accuracy of such system is still a issue due to variety of writing patterns, size, slant, ink, and writing style. So In this paper, a novel approach for Devanagari handwritten numerals recognition based on global and local structural features is proposed. Probabilistic NeuralNetwork (PNN) Classifier is used to classify the Devanagari numerals separately.

Keywords: Character Recognition, Off-line Handwriting Recognition, Segmentation, Feature Extraction, Training and Recognition and probabilistic Neural network.

Introduction

Handwritten numeral recognition is an integral part of the handwritten character recognition system. The problem of the handwritten numeral recognition is a complex task due to the variations among the writers such as style of writing, shape, stroke etc., and it has variety of applications in various fields like reading postal zip code, passport number, employee code, bank cheque, and form processing.

For recognition of image object one need to extract the potential features. Thus, Feature extraction plays a vital role in Pattern Recognition and Image Processing systems in general and character recognition systems in particular. Ivind and Jain [1] present a survey of various feature extraction methods used in character recognition system.

The problem of numeral recognition has been studied for decades and many methods have been proposed, e.g. template matching, dynamic programming, hidden Markov modeling, neural network, expert system and combinations of these techniques [2, 3, and 4]. Recognition of character/numeral in foreign languages like English, Chinese, Japanese, and Arabic reported many authors.

The paper is organized as follows: Section 2 of the paper contains the Related Works .The Classification method and its algorithm is the subject matter of Section 3. Section 4 contains the conclusion part of the problem. Section 5 and 6 fallows Acknowledgment and References.

Related Works

Historically, handwritten character recognition systems have evolved in three ages:
1900-1980 Early ages-- The history of character recognition can be traced as early as 1900, when the Russian Scientist Trying attempted to develop an aid for visually handicapped. The first character recognizers appeared in the middle of the 1940s with the development of the digital computers. The early work on the automatic recognition of characters has been concentrated either upon machine printed text or upon small set of well distinguished handwritten text or symbols. Machine-printed CR systems in this period generally used template matching in which an image is compared to a library of images. For handwritten text, low level image processing techniques have been used on the binary image to extract feature vectors, which are then fed to statistical classifiers. Successful, but constrained algorithms have been implemented mostly for Latin characters and numerals. However, some studies on Japanese, Chinese, Hebrew, Indian, Cyrillic, Greek

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 1.852

and Arabic characters and numerals in both machine-printed and handwritten cases were also initiated [5]. The commercial character recognizers were available in 1950s, when electronic tablets capturing the x-y coordinate data of pen-tip movement was first introduced. This innovation enabled the researchers to work on the on-line handwriting recognition problem. A good source of references for on-line recognition until 1980 can be found in [6].

1980-1990 Developments-- The studies until 1980 suffered from the lack of powerful computer hardware and data acquisition devices. With the explosion on the information technology, the previously developed methodologies found a very fertile environment for rapid growth in many application areas, as well as CR system development [7]. Structural approaches were initiated in many systems in addition to the statistical methods. These systems broke the character image into a set of pattern primitives such as lines and curves. The rules were then determined which character most likely matched the extracted primitives. However, the CR research was focused on basically the shape recognition techniques without using any semantic information. This led to an upper limit in the recognition rate, which was not sufficient in many practical applications. Historical review of CR research and development during this period can be found in [8] and [9] for off-line and on-line case, respectively.

After 1990 Advancements-- The real progress on CR systems is achieved during this period, using the new development tools and methodologies, which are empowered by the continuously growing information technologies. In the early nineties, Image Processing and Pattern Recognition techniques are efficiently combined with the Artificial Intelligence methodologies. Researchers developed complex CR algorithms, which receive high-resolution input data and require extensive number crunching in the implementation phase. Nowadays, in addition to the more powerful computers and more accurate electronic equipments such as scanners, cameras and electronic tablets, we have efficient, modern use of methodologies such as Neural Networks, Hidden Markov Models, Fuzzy Set Reasoning and Natural Language Processing. The recent systems for the machine-printed off-line [10] and limited vocabulary, user dependent on-line handwritten characters [11] are quite satisfactory for restricted applications. However, there is still a long way to go in order to reach the ultimate goal of machine simulation of fluent human reading, especially for unconstrained on-line and off-line handwriting.

Devanagri Script

India is a multi-lingual and multi-script country comprising of eighteen official languages. One of the defining aspects of Indian script is the repertoire of sounds it has to support. Because there is typically a letter for each of the phonemes in Indian languages, the alphabet set tends to be quite large. Most of the Indian languages originated from Bramhi script. These scripts are used for two distinct major linguistic groups, Indo-European languages in the north, and Dravidian languages in the south .

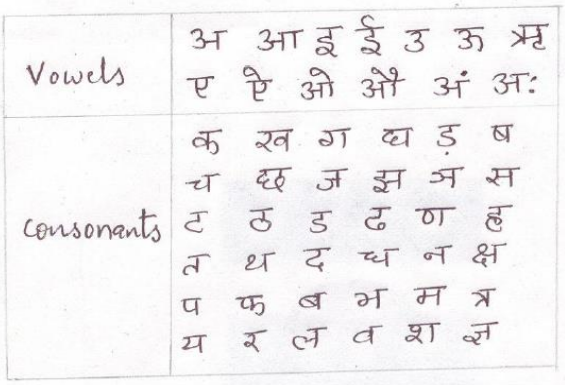

Figure 1 :Devnagari character

Devnagari is the most popular script in India. It has 11 vowels and 33 consonants. They are called basic characters. Vowels can be written as independent letters, or by using a variety of diacritical marks which are written above, below, before or after the consonant they belong to. When vowels are written in this way they are known as *modifiers* and the characters so formed are called *conjuncts*. Sometimes two or more consonants can combine and take new shapes. These new shape clusters are known as *compound characters*.

Theoretical Methods

Neural network :

To develop an accurate OCR system is a complicated task and requires a lot of effort. Such types of systems are usually complicated and can hide a lot of logic behind the code. To achieve the good performance and improved quality of recognition, the artificial neural network in OCR applications performs the important function. Another benefit of using neural network in OCR is extensibility of the system ability to recognize more character sets than initially defined. There are the different types of neural networks which can be used

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 1.852

for the character recognition. The artificial Neural Network is a computing architecture which consists of 'neural' processors connected in a parallel sequence.

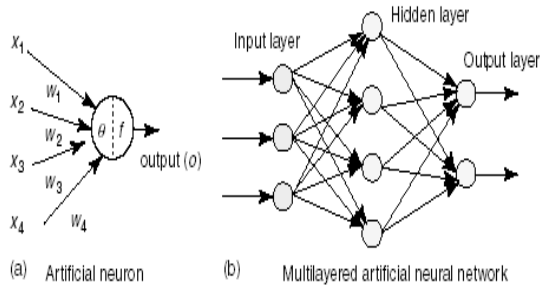

Figure 2: (a) Artificial Neural Network , (b) Multilayer ANN

Artificial Neural Networks can perform the computation at a higher rate as compared to the classical techniques because it has the parallel nature. The output from one node goes to another node as the input and the final decision depends on the complex interaction of all nodes. The neural network can be categorized in to two types, feed forward and feedback network. The feedback network is also known as the recurrent network. In OCR system, the most common neural network is multilayer perceptron (MLP) network which is of type feed forward network. The interesting Feature of MLP is that it provides the confidence in the character classification. First of all MLP is proposed by U. Bhattacharya et al [12]. M. Egmont- Petersen has shown the comparison of various NN classifiers like Feed forward, Neuro-fuzzy system etc. for English language, K.Y Rajput et al[13] have also used classifier like back propagation type which is based on Genetic algorithm and also classification along with fusion of NN and Fuzzy logic.

Proposed Method

A. System Design:

Preprocessing includes the steps that are necessary to bring the input data into an acceptable form for feature extraction. The raw data, depending on the data acquisition type, is subjected to a number of preliminary processing stages.

Preprocessing stage involves noise reduction, slant correction, size normalization and thinning. Among these size normalization and thinning are very important. Normalization is required as the size of the numeral varies from person to person and even with the same person from time to time.

Thinning provides a tremendous reduction in data size, thinning extracts the shape information of the characters. It can be considered as conversion of off-line handwriting to almost on-line data. Thinning is the process of reducing thickness of each line of pattern to just a single pixel. The reduced pattern is known as the skeleton and is close to the medial axes, which preserves the topology of the image.

For classification and recognition nearest neighbor classifier and feed forward back propagation neural network classifiers are used.

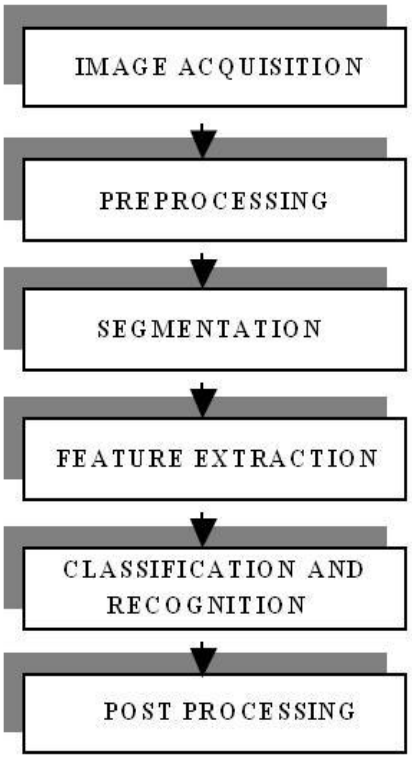

Figure 2 :Block diagram of handwritten neural recognition system

Steps involved in implementation of proposed method are given as following and Shown in Figure.2 Block diagram of handwritten neural recognition system.

B. Algorithm:

Step 1: Compute the input image centroid

Step 2: Divide the input image in to **n** equal zones.

Step 3: Compute the distance between the image centroid to each pixel present in the zone.

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 1.852

Step 4: Repeat step 3 for the entire pixel present in the zone.

Step 5: Compute average distance between these points.

Step 6: Repeat this procedure sequentially for the entire zone.

Step 7: Finally, **n** such features will be obtained for classification and recognition

Conclusion

The important applications of pattern recognition include Character recognition technique. Character recognition of Indian scripts is in its preliminary stage and a lot of research is needed to handle the complexity and issues in Devanagari character recognition (DCR). The accurate recognition is directly depending on the nature of the material to be read and by its quality. The recognition process needs to be much efficient and accurate to recognize the characters written by different users.

This system involves implementation of Devanagari Handwritten Numeral recognition which includes image acquisition, preprocessing, segmentation, feature extraction, classification and recognition and at last post-processing.

References

[1] *Ivind due trier, anil Jain, torfiinn Taxt, "A feature extraction method for character recognition-A survey ", pattern Recg, vol 29, No 4, pp-641-662, 1996*

[2] *A.L.Koerich, R. Sabourin, C.Y.Suen, "Large offlineHandwritten Recognition: A survey", Pattern Analysis Application, 97-121, 2003.*

[3] *J.D. Tubes, A note on binary template matching. Pattern Recognition, 22(4):359-365, 1989.*

[4] *B.V. Dhandra, R.G.Benne and Mallikargun Hangargi, "Isolated Handwritten Devnagari Numeral recognition based on Template matching", IEEE-ACVIT -07, pp.1276-1282,Dec-2007.*

[5] *S. Mori, H. Nishida, H. Yamada, Optical Character Recognition, Wiley, 1999.*

[6] *S. Mori, K. Yamamoto, M. Yasuda, "Research on Machine Recognition Of Hand printed Characters", IEEE Trans. Pattern Analysis, Machine Intelligence, vol.6, no.4, pp.386-404, 1984.*

[7] *C. Y. Suen, C. C. Tappert, T. Wakahara, "The State of the Art in on-Line Handwriting Recognition", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 12, no. 8, Pp.787-808, 1990.*

[8] *V. K. Govindan, A. P. Shivaprasad, "Character ecognition- A review", Pattern Recognition vol.23, no.7, Pp.671- 683, 1990.*

[9] *T. Steinherz, E. Rivlin, N. Intrator, "Off-line Cursive script Word Recognition - A survey", Int. Journal Document Analysis and Recognition, vol.2, no.2, pp.90- 110, 1999*

[10] *C. Y. Suen, C. C. Tappert, T. Wakahara, "The State of the Art in On-Line Handwriting Recognition", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 12, no. 8, pp.787- 808, 1990.*

[11] *I. Bazzi, R. Schwartz, J. Makhoul, "An Omnifont Open Vocabulary OCR System for English and Arabic", IEEE Trans. Pattern Analysis and Machine Intelligence, vol.21, no.6, pp.495-504, 1999.*

[12] *Sandhya Arora et al., "Performance Comparison of SVM and ANN for Handwritten Devnagari Character Recognition" IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 3, May 2010.*

[13] *K. Y. Rajput and Sangeeta Mishra, "Recognition and Editing of Devnagari Handwriting Using Neural Network", SPITIEEE Colloquium and Intl. Conference, Mumbai, India.*