International Journal of Soft Computing and Engineering (IJSCE)
ISSN: 2231-2307, Volume-2, Issue-4, April 2013

Devnagari Script Character Recognition Using Genetic Algorithm for Get Better Efficiency

Vedgupt Saraf, D.S. Rao

***Abstract*—Character recognition is the mechanical or electronic translation of scanned images of handwritten, typewritten or printed text into machine-encoded text. In India, more than 300 million people use Devanagari script for documentation. There has been a significant improvement in the research related to the recognition of printed as well as handwritten Devanagari text in the past few years. The problem arises in Devnagari script character recognition using quadratic classifier provides less correctness and less efficiency. For the answer of the above problem and for get better efficiency we use the genetic algorithm. It will give the better results from the above methods.**

The idea of genetic algorithm comes from the fact that it can be used as an outstanding means of combining various styles of writing a character and generates new styles. Closely observing the ability of human mind in the recognition of handwriting, we find that humans are able to recognize characters even though they might be seeing that style for the first time. This is possible because of their power to visualize parts of the known styles into the unknown character. We try to represent the same power into the machines.

***Index Terms*-Handwritten Character Recognition, On-line and Off-line Character Recognition, Genetic Algorithms, Segmentation.**

I. INTRODUCTION

Character recognition is considered as one of the important technology for today's world and it is used in many fields such as artificial intelligence, computer vision, pattern matching etc. There are two types of character recognition systems.

1. Optical character recognition: - It is also considered as offline character recognition. In this type of character recognition either handwritten, type written or printed text is converted into digital format. It does not have the advantage of recognizing direction of the movements while writing the character.

2. Intelligent character recognition: - It is also considered as online character recognition. It recognizes character on the basis of the direction of the motion while writing character. This method is generally available on touchpad, touch screen cell phones etc.

Manuscript received on April, 2013.

Vedgupt Saraf, Computer Science and Engineering Department, Indore Institute of Science and Technology, INDORE, INDIA.

Dr. D.S. Rao, Computer Science and Engineering Department, Indore Institute of Science and Technology, INDORE, INDIA.

The problem of handwriting recognition can be classified into two main groups, off-line and on-line recognition, according to the format of handwriting inputs. In offline recognition, only the image of the handwriting is available, while in the on-line case temporal information such as pen tip coordinates, as a function of time, is also available. Many applications require off-line HWR capabilities such as bank processing, mail sorting, document archiving, commercial form-reading, office automation, etc. So far, off-line HWR remains an open problem, in spite of a dramatic boost of research in this field and the latest improvement in recognition methodologies.

II. HISTORY

A. *1900–1980 Early Ages:* The history of CR can be traced as early as 1900, when the Russian scientist Tyuring attempted to develop an aid for the visually handicapped. The first character recognizers appeared in the middle of the 1940s with the development of digital computers.

B. *1980–1990 Developments:* Studies up until 1980 suffered from the lack of powerful computer hardware and data acquisition devices. With the explosion of information technology, the previously developed methodologies found a very fertile environment for rapid growth in many application areas, as well as CR system development. Structural approaches were initiated in many systems in addition to the statistical methods.

C. *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.

Nowadays Evolutionary Algorithm (EA) has been successfully applied to find the answer of numerous problems from pattern recognition area. It uses biological evolution viz. reproduction, mutation, recombination and selection. The generally used Evolutionary Algorithms are Genetic Algorithm, Evolutionary Programming, Evolutionary Strategy, Genetic Programming, Particle Swarm Optimization, Artificial Immune, Ant Colony Optimization and Invasive Weed Optimization and Bee's Optimization.

Handwriting recognition has always been a special problem. The problem increases when we operate it in the offline mode. We see a lot of work has been done in this area in the past few years. The solutions being proposed mainly use Artificial Neural Networks (ANN) and Hidden Markov Models (HMM) for solving the problem. Genetic algorithms have not been applied much. They have been applied for feature selection optimization.

III. METHODOLOGIES OF CR SYSTEM

In this section the available methodologies to develop the stages of the CR system are presented. A general character recognition system has different phases are as given below:

1. Data acquisition

2. Pre-processing
3. Segmentation
4. Feature extraction
5. Classification
6. Post-processing

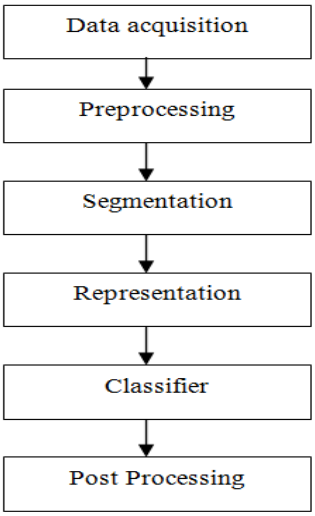


Fig.1. Typical Flowchart of CR Methodology

A. Data acquisition

The development in automatic character recognition systems is evolved in two categories according to the form of data gaining:

- On-line character recognition systems
- Off-line character recognition systems

Off-line character recognition capture the data from paper through optical scanners or cameras where the on-line recognition systems use the digitizers which directly capture writing with the order of the strokes, speed, pen- up and pen-down information.

B. Preprocessing

The raw data, depending on the data acquisition type, is subjected to a number of preliminary processing steps to make it usable in the descriptive stages of character analysis. Preprocessing aims to produce data that are easy for the CR systems to operate accurately. The main objectives of preprocessing are:-

1) *Noise Reduction:* The noise, introduced by the optical scanning device or the writing instrument, causes disconnected line segments, bumps and gaps in lines, filled loops, etc. The distortion, including local variations, rounding of corners, dilation, and erosion, is also a problem. Prior to the CR, it is necessary to eliminate these imperfections. Hundreds of available noise reduction techniques can be categorized in three major groups.

a) Filtering
b) Morphological operations
c) Noise modeling

2) Normalization of the data: Normalization methods aim to remove the variations of the writing and obtain standardized data. The following are the basic methods for normalization.

a) Skew Normalization and Baseline Extraction
b) Slant Normalization
c) Size Normalization
d) Contour Smoothing

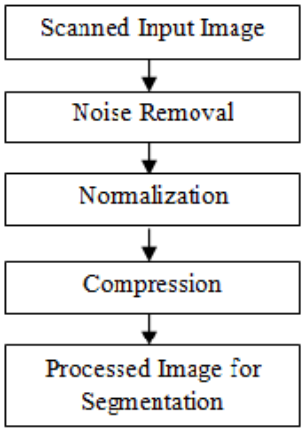


Fig.2. Flowchart for Pre-processing

3) *Compression:* Space domain techniques are necessary for compression. Two important techniques are thresholding and thinning. Thresholding reduces the storage requirements and increases the rate of processing by converting the gray-scale or color images to binary image by taking a threshold value. Thinning extracts the shape information of the characters.

a) Thresholding
b) Thining

C. Segmentation

Segmentation is a significant stage in CR system because it affects the rate of recognition. Segmentation can be external and internal. External segmentation is the separation of various writing units, such as paragraphs, sentences or words. In internal segmentation an image of series of characters is decomposed into sub-images of individual character. There are two types of segmentation: external segmentation, which is the isolation of various writing units, such as paragraphs, sentences, or words, and internal segmentation, which is the isolation of letters, especially in cursively written words.

a) External segmentation
b) Internal segmentation

D. Representation

The feature extraction step selects and prepares data which is used by a classifier to get the recognition task. Feature extraction involves representing a handwriting text by a set of discriminative features. The feature representation is based on removal of certain types of information from the image.

E. Classification

The classification phase is the decision making part of the recognition system. The performance of a classifier relies on the quality of the features. There are many existing Classical and soft computing techniques for handwriting recognition. They are given as:

1) Template matching
2) Statistical techniques
3) Structural techniques
4) Neural networks (NNs)
5) Fuzzy- logic technique
6) Evolutionary computing techniques

F. Post processing

Post-processing stage is the last stage of the proposed recognition system. It prints the corresponding recognized characters in the structured text form.

IV. GENETIC ALGORITHM

A GAs is an optimization and search method utilized in computer science to find fairly accurate solutions to problems. It is inspired by processes in biological evolution such as natural selection, inheritance, recombination, and mutation. 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 and 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.

V. DEVNAGARI LANGUAGE

In India, more than 300 million people use Devanagari script for documentation. There has been a significant improvement in the research related to the recognition of printed as well as handwritten Devanagari text in the past few years.

Devnagari is the most accepted script in India and the most popular Indian language Hindi is written in Devnagari script. Nepali, Sanskrit and Marathi are also written in Devnagari script. Moreover, Hindi is the national language of India and the third mainly popular language in the world. So, the work on Devnagari script is very helpful for the country.

The alphabet of present Hindi consists of 14 vowels and 33 consonants. These characters may be called basic characters. Writing style in Devnagari script is from left to right. The concept of upper/lower case is absent in Devnagari script.

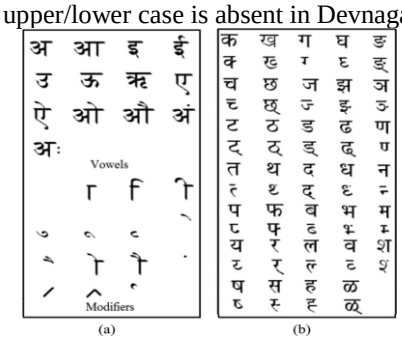


Fig.3. Printed samples of Devnagari characters considered in this experiment (a) Vowels (b) Consonants.

VI. GENERAL PROCEDURE

Handwriting recognition is a famous problem which involves the recognition of whatever input is given in form of image, scanned paper, etc. The handwriting recognition generally involves the following steps:-

A. *Segmentation:* This step deals with the breaking of the lines, words and finally getting all the characters separated. This step involves the identification of the boundaries of the character and separating them for further processing. In this algorithm we assume that this step is already done. Hence the input to our system is a single character.

B. *Preprocessing:* This step involves the initial processing of the image, so that it can be used as an input for the recognition system. In this algorithm we assume that a part of this step has been done. We assume that the character segmented is made thin to a unit pixel thickness. Various algorithms may be used for this purpose. The further processing is done by our algorithm.

C. *Recognition:* Once the input image is available in good condition, it may be processed for recognition. The role of the recognition system is to identify the character. Our algorithm uses an image as an input for the same.

VII. GENETIC ALGORITHM APPLYING

When the features of the characters in the sub-word are determined, the next phase is to recognize the characters of the sub-word. The genetic algorithm approach will be used for this purpose.

Firstly we retrieve the image from the database then we segmented that image into lines. After segmented lines into words and then segment words into sub-words. Now we normalized the image and after determine the number of peaks in that image. Then we detect loop in the peak and after that determine the complementary character. Now we compute the height and width of the peak and determine left and right connection. After it send this peak's string to the genetic algorithm. Now we apply condition that if find last peak in sub-word then go to the next condition which is last sub-word in word exist then go to the next condition which is last word and if last word is found then end the algorithm and we recognized the characters.

In genetic algorithm we have initial population and then we applying three operators in which first is selection which selects the strings and then we apply crossover operator which recombining those selected strings and after that we apply mutation operator those changes strings of 0's and 1's form. Now we apply condition and check optimization criteria met or not, if met then selects the best string which is our solution and if not then send it to in the initial population. This is the process of genetic algorithm which is also given in flow chart below.

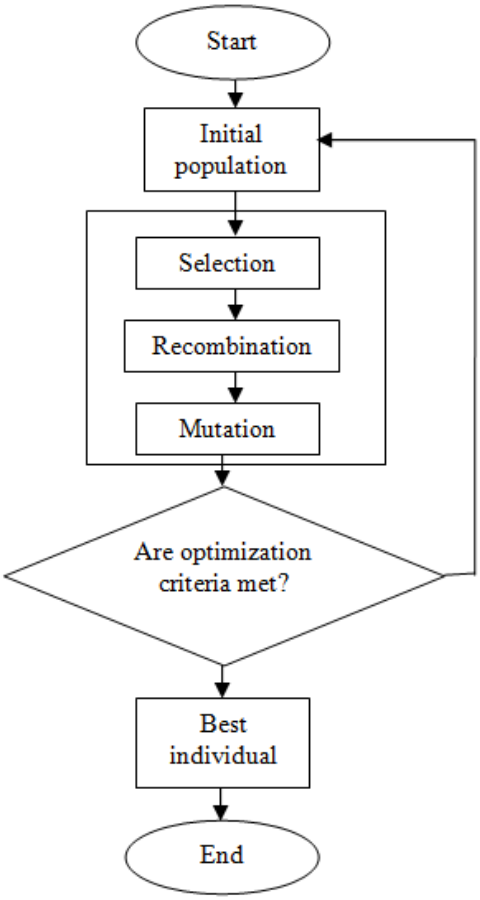


Fig.4. Flow Chart of Genetic Algorithm

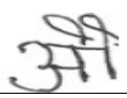
VIII. RESULT AND COMPARISON

Data used for the present work were collected from different individuals.

TABLE I. INDIVIDUAL ACCURACY OF SOME DEVNAGARI CHARACTERS

Character	Accuracy	Character	Accuracy
	97.84%		98.52%
	98.08%		97.98%

TABLE II. MAIN CONFUSING PAIRS OF DEVNAGARI CHARACTERS

Confusing character pairs		Confusion rate (Computed from overall samples)
		0.25%
		0.17%

IX. CONCLUSION AND FUTURE SCOPE

India is a multi-lingual and multi-script country comprises of eleven different scripts and not much work has been done towards off-line handwriting recognition of Indian scripts. In this paper we present a genetic algorithm scheme towards the recognition of off-line Devnagari handwritten characters. We tested our proposed system on different individuals' samples and obtained 98.78% recognition accuracy.

This problem could have been solved with the absence of Genetic Algorithms as well. In order to see the importance of Genetic Algorithms in the problem, we tested the data in the absence of the application of genetic algorithms. We found that the genetic algorithms were useful in the following manner. We hope this work will be helpful to the researchers for the work towards other Indian script characters.

REFERENCES

[1] Shabana Mehfuz1, Gauri Katiyar2, "Intelligent Systems for Off-Line Handwritten Character Recognition: A Review ", International Journal of Emerging Technology and Advanced Engineering, Volume 2, Issue 4, April 2012.

[2] Shabana Mehfuz1, Gauri Katiyar2, "Intelligent Systems for Off-Line Handwritten Character Recognition: A Review ", International Journal of Emerging Technology and Advanced Engineering, Volume 2, Issue 4, April 2012.

[3] Prof. Swapna Borde, Ms. Ekta Shah, Ms. Priti Rawat, MsVinaya Patil, "Fuzzy based handwritten character recognition system" , *National Conference on Emerging Trends in Engineering & Technology (VNCET-30 Mar'12)*

[4] U. Pal1, N. Sharma1, T. Wakabayashi2 and F. Kimura2, "Off-Line Handwritten Character Recognition of Devnagari Script", Ninth International Conference on Document Analysis and Recognition (ICDAR 2007).

[5] Rahul KALA1, Harsh VAZIRANI2, Anupam SHUKLA3 and Ritu TIWARI4, "Offline Handwriting Recognition using Genetic Algorithm", IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 2, No 1, March 2010.

[6] Nafiz Arica, Student Member, IEEE and Fatos T. Yarman-Vural, Senior Member, IEEE, "An Overview Of Character Recognition Focused On Off-line Handwriting", Computer Engineering Department, Middle East Technical University, Ankara, Turkey, Manuscript received June21,1999.