Impact Factor (SJIF): 3.632

International Journal of Advance Research in Engineering, Science & Technology

e-ISSN: 2393-9877, p-ISSN: 2394-2444

Volume 3, Issue 7, July-2016

Recognition of English Handwritten Characters and Digits using Genetic Algorithm

Hemant Pandey[1], Sanjeev Ghosh [2]

[1] Electronics & Telecommunication, *SLRCTE*
[2] Electronics & Telecommunication, *TCET*

Abstract *—Many researchers have worked on recognition of printed characters, numerals but a very little amount of work is available on handwritten character recognition. Handwritten Character Recognition is one of the most complex and tricky field in the field of pattern recognition. Variety of works has been done for English languages having various recognition rates because of different shape and writing styles of individuals. There are enumerable styles in which a character may be written. 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, it is found 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. In the proposed work, the same power is represent into the machines by using genetic algorithm.*

Keywords: character recognition, genetic algorithm, offline handwritten character, morphological operation, erosion and dilation, uppercase and lowercase.

I. INTRODUCTION

Character recognition is process of detecting and recognizing characters from input image and convert it into text [2]. More precisely it is a process of detecting, segmenting and identifying characters from image. It contributes immensely to the advancement of automation process and improving the interface between man and machine in many applications.

Character recognition is a process, which associates a symbolic meaning with objects (letters, symbols and numbers) drawn on an image, *i.e.,* character recognition techniques associate a symbolic identity with the image of a character [17]. Mainly, character recognition system takes the raw data that further implements the process of preprocessing of any recognition system. On the basis of that data acquisition process, character recognition system can be classified into following categories: -

Character recognition can be categorized into following two parts: -

1. Online Character Recognition
2. Offline Character Recognition

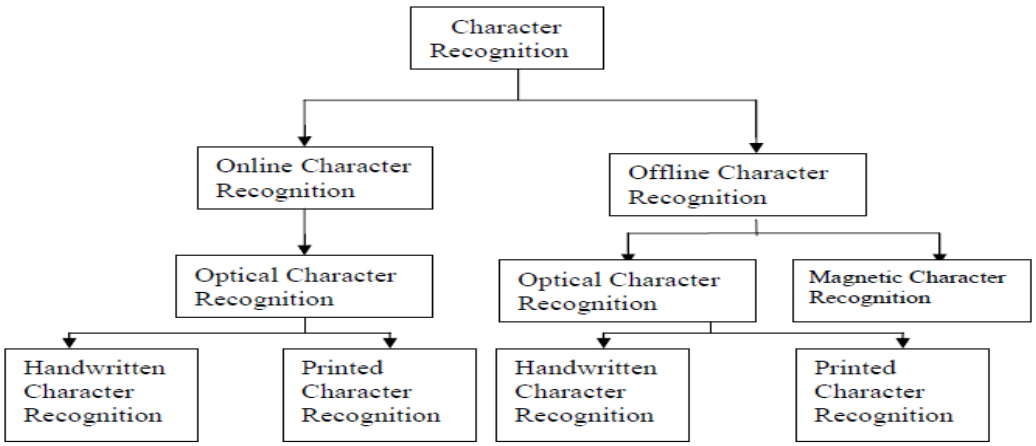

Figure 1. Classification of Character Recognition

Off-line handwriting recognition refers to the process of recognizing words that have been scanned from a surface (such as a sheet of paper) and are stored digitally in grey scale format. After being stored, it is conventional to perform further processing to allow superior recognition. In case of online handwritten character recognition, the handwriting is captured and stored in digital form via different means. Usually, a special pen is used in conjunction with an electronic surface [17]. As the pen moves across the surface, the two- dimensional coordinates of successive points are represented as a function of time and are stored in order. It is generally accepted that the on-line method of recognizing handwritten text has achieved better results than its off-line counterpart [7]. This may be attributed to the fact that more information may be captured in the on-line case such as the direction, speed and the order of strokes of the handwriting.

II. RELATED WORK

The Handwritten recognition for different languages like Thai, Chinese, Sanskrit English, Devnagri etc have been used using genetic algorithms. Although, the languages are different but the basic steps for recognition are the same i.e., preprocessing, feature extraction and genetic algorithm for the recognition. But the different researchers have used different segmentation techniques and different feature extraction methods and hence got different accuracy. This paper aims do combine the best segmentation technique and feature extraction method to get a better efficiency for the recognition of English characters and digits.

Euclidian distance method and Mahalanobis distance method are used as a fitness function for the recognition of Kannada handwritten characters. The result of the system is quite poor for both the fitness functions [1]. For the Handwritten Sanskrit word recognition, Prewitt's operator is used for the edge detection for the recognition. Freeman chain code is used as the representation technique of an image character [2].

The combination of genetic algorithm and neural network is used for the recognition of handwritten mathematical symbol [3]. This combination providesefficient, parallel, and strong robustness and it solve complex optimization problems as genetic algorithm is used and neural network gives ability of self-evolution and self-adaption. But recognition speed is slow and also it does not recognize the complex mathematical symbols. The accuracy of the system is 90.6%.

For the recognition of English handwritten characters, graph theory is used. [4]. The pool of images of characters is created which is then converted into graph. The graph of every character was intermixed to generate styles intermediate between the styles of parent character. Character recognition involves the matching of the graph generated from the unknown character image with the graphs generated by mixing. The only disadvantage is the input is taken as a single character.

Hence, the handwritten recognition system needs good accuracy with good recognition speed. Not much work is done on lowercase English characters and most work done involves only uppercase English characters. Also, the systems described in literature above require only individual characters or digits as input. Systems that can accept whole sentences or paragraphs are desired with high accuracy and efficiency.

III. SYSTEM DESIGN

The approach has been divided into two parts, first to create database and second to actually recognize the characters from the image.

3.1 Creation of database

In database creation, 55 samples for each number (0 to 9), alphabet uppercase (A to Z) and lowercase (a to z) has been considered. The entire procedure of database creation is shown in figure 2. First, the image is converted to grayscale image and then converted to binary image. After that it is filtered with the help of median filtering. Then each image is resized to resolution 100X50 to keep the size of database limited and to have sufficient speed of operation and saved to database. Total images saved in the database are 3410.

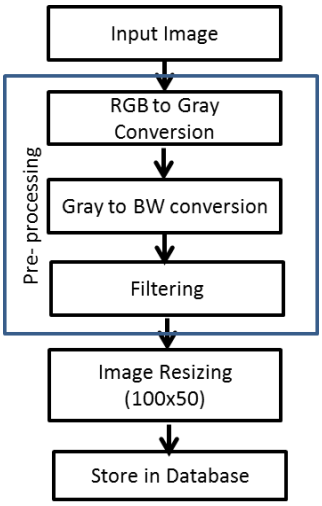

Figure2. Database Creation

3.2 character recognition technique

In this process, the image containing handwritten characters has been given as input image to the system. The character recognition technique is shown in Figure 3. The image is preprocessed that is it undergoes RGB to gray conversion, followed by binary conversion and median filtering. Morphological operation like dilation and erosion are used in preprocessing method to simplifies an image by removing irrelevant details and errors with which the essential characteristics of the form of objects remain intact.

Then from the processed image the first line is identified by the examining the connected pixel components in the given rows. This line is separated from the remaining image. Now this line is considered as an input. After this, again by checking the relation in connected component one word is separated from all other words and then it is further separated into letters and then each letter (character) is considered as an input image. Now this image resized to 100x50 same as that of images in database. Then the image is compared with all the images in the database one by one and then matching is done with the help of Genetic algorithm. The GA works to find the index of best possible match from the database. From that index the actual word is fetched and writes to text file. The same procedure is done for the remaining words in the line. After the end of line the next line is considered and each word from the line is examined and recognized one by one. After the end of all lines, the process is completed and the output is displayed in the text file.

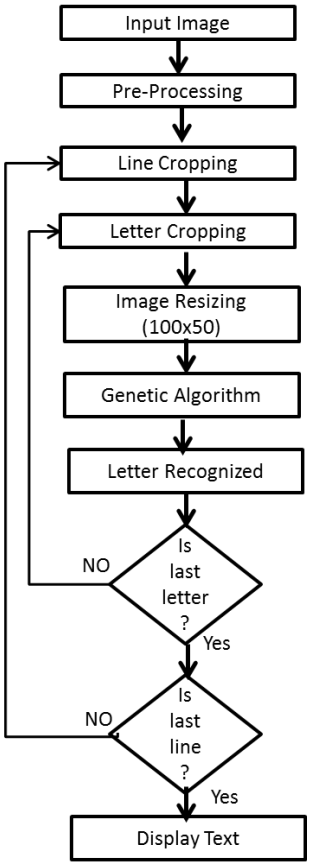

Figure3. Character Recognition Technique

IV. GENETIC ALGORITHM

A GAs is an optimization and search method utilized in computer science to find fairly accurate solutions to problems [6]. 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.

4.1 The Basic Genetic Algorithm [7]:

- Start with a large "population" of randomly generated "feasible solutions" to a problem
- Repeatedly do the following:
  - Evaluate each of the feasible solution
  - Keep a subset of these solutions (the "best" ones)
  - Use these solutions to make generation of new population
- Quit when you have a satisfactory solution (or you run out of time)

V. RESULT

Each English character as well as digit is individually tested to get the accuracy of particular character and digit. Each character is test on different types of writing styles by using ball pens and marker. The accuracy of different characters are found to be different because of their different shapes.Table below show the recognition rate of the digits, upper case and lower case.

Character	Recognition %	Character	Recognition %
A	80	N	85
B	50	O	50
C	75	P	59
D	63	Q	80
E	50	R	80
F	92	S	55
G	80	T	88
H	70	U	71
I	50	V	84
J	80	W	75
K	70	X	55
L	100	Y	88
M	80	Z	96

Table1. Recognition rate of the uppercase

Character	Recognition %	Character	Recognition %
a	50	n	90
b	87	o	84
c	75	p	80
d	67	q	75
e	50	r	60
f	59	s	71
g	90	t	88
h	60	u	71
i	10	v	67
j	70	w	92
k	88	x	75
l	90	y	80
m	90	z	67

Table2. .Recognition rate of the lower case

Digits	Recognition %
0	25
1	55
2	88
3	53
4	53
5	85
6	30
7	63

8	86
9	84

Table3. Recognition rate of the digits

When average accuracy is calculated for uppercase, lower case and digits, it is found that for uppercase it is 73.33%, for lower case it is 71.5 % and for digit it is 62.2%. Also the characters are tested in groups (words) and obtained 91.98% recognition accuracy.

VI. CONCLUSION

Handwritten English character recognition using Genetic algorithm is discussed here. It has been found that recognition of handwritten English characters is a very difficult task. Following are main reasons for difficulty in recognition of English characters:

- Some English characters are similar in shape (for example digit "0", letter "O" and "D").
- Different, or even the same person can write in different style at different times, depending on the types of pen marker and pencil, the width of the line, the slight rotation of the paper, the type of paper and the mood and stress level of the person.
- The characters can be written at different location on paper or in window
- Characters can be written in different fonts.

VII. FUTURE SCOPE

The recognition accuracy using genetic algorithm proposed here can be further improved. The number of character set used for training is reasonably low and the accuracy of the system can be increased by taking more training character sets.

Still, there are some facts which can be improved like the proper spacing between the words and paragraphs. Also, the characters written on the pages have a horizontal line which is difficult for the system to recognize it first and then remove it during the output. Introduction of special characters of English (! @ # " ; ? ,) can improve the efficiency of the system to a very great extent.

REFERENCES

[1] Vijaya S. Waghmare and Leena R. Ragha, "Application of Genetic Algorithm for Handwritten Kannada Characters" 2013 3rd IEEE International Advance Computing Conference (IACC).

[2] NamitaDwivedi, Kamal Srivastava and NeelamArya. "Sanskrit Word Recognition Using Prewitt's Operator And Support Vector Classification", 2013 IEEE International Conference on Emerging Trends in Computing.

[3] Cao Xinyan, Yu Hongli,WangXin, " Handwritten Mathematical Symbol Recognition Based on Niche Genetic Algorithm", 2013 Third International Conference on Intelligent System Design and Engineering Applications.

[4] Rahul Kala, Harsh Vazirani, AnupamShukla And RituTiwari, "Offline Handwriting Recognition using Genetic Algorithm", IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 2, No 1, March 201

[5] ChomtipPornpanomchai, MontriDaveloh, "Printed Thai Character Recognition By Genetic Algorithm",Proceedings of the Sixth International Conference on Machine Learning and Cybernetics, Hong Kong, 19-22 August.

[6] VedguptSaraf, D.S. Rao," Devnagari Script Character Recognition Using Genetic Algorithm for Get Better Efficiency", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-4, April 2013.

[7] Shilpa Jumanal, Ganga Holi, "On-line Handwritten English Character Recognition Using Genetic Algorithm", International Journal of Computer Trends and Technology (IJCTT) – volume 4 Issue 6–June 2013.

[8] Li-Xin Zhang, Yan-Nan Zhao, Ze-Hong Yang, Jia-Xin Wang, "Feature Selection In Recognition Of Characters" Proceedings of the First International Conference on Machine Laming and Cybernetics, Beijing, 4-5 November 2002.

[9] B. Bontempi, A. Marcelli, "A genetic learning system for on-line character recognition in Pattern Recognition," 1994. Vol. 2-Conference B: Computer Vision & Image Processing., Proceedings of the 12th IAPR International. Conference on vol. 2, IEEE, 1994.

[10] ChihiroIga and Toru Wakahara," Character Image Reconstruction from a Feature Space Using Shape Morphing and Genetic Algorithms," Proceedings of the 9th Int'l Workshop On Frontiers in Handwriting Recognition, 2004 IEEE

[11] P. Vijayalakshmi, M.Sumathi. "Design of Algorithm for Vehicle Identification by Number Plate Recognition," IEEE- Fourth International Conference on Advanced Computing, ICoAC 2012 MIT, Anna University, Chennai. December 13-15, 2012

[12] R. Sharapov, Genetic Algorithms: Basic Ideas, Variants and Analysis. na, 2007.

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

[14] 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).

[15] Mengjie Zhang, Victor Ciesielski, Genetic Programming forMultiple Class Object Detection, Springer Berlin / Heidelberg,ISBN No. 978-3-540-66822-0, 2007.

[16] HananAljuaid (2012). Applying Genetic Algorithm in Multi Language'sCharactersRecognition, Genetic Algorithms in Applications, Dr. RustemPopa (Ed.), ISBN: 978-953-51-0400-1

[17] Naveen Garg Handwritten Gurumukhi Character Recognition Using Neural NetworksComputer Science And Engineering Department Thapar University Patiala – 147004, June- 2009.