International Journal of Ethics in Engineering & Management Education
Website: www.ijeee.in (ISSN: 2348-4748, Volume 5, Issue 7, July 2018)

OFFLINE HANDWRITTEN CHARACTER RECOGNITION OF SUGALI SCRIPT

Anuma Sharma[1], Dr.Rakesh K Bhujade, Dr.Amit Sinhal[3]
[1]M.Tech Scholar, Department of Information Technology, TIT Bhopal, madhuramjain0@gmail.com,India;
[3]Assistant Professor,[3]Head of Department,Department of Information Technology, TIT Bhopal,

Abstract: **Handwritten character recognition (HCR) is a border area of research in field of character recognition and image processing and has been investigated under the frame work of a character recognition (OCR) and pattern recognition. In this study we propose a solution for performing character recognition in Sugali language, an Indo –Aryan language spoken in Andhra Pradesh and a number of states including MP, HP, Gujarat, TN, Maharashtra, Karnataka, Orissa and WB. This language is also known as Lambadi or Lamani. These characters are further more specially recognized by pattern recognition using neural networks. There are some different techniques of OCR systems such as optical scanning, location, segmentation, pre-processing, segmentation, representation, feature extraction and recognition and post processing.**

Keywords: **Handwritten, OCR, PNN, RBF, GA**

1. INTRODUCTION

OCRis aprocess of automatic recognition of characters from optically scanned. OCR is a process of classification of an optical patterns contain in a digital image corresponding to an alphanumeric or some different characters. The pattern recognition is the process to classify many areas such as a computer vision, face recognition, speech recognition, character recognition, signal classification and analysis, medical diagnosis etc. This study focuses mainly on offline handwritten character recognition of indo –Aryan language namely Sugali language. OCR is one in all the foremost fascinating and difficult areas of pattern recognition with varied sensible applications. I also contribute the advancement of an automation process and can improve the interface between man and machine in much application. The character recognition is broadly categorized into optical character recognition (OCR) and handwritten character recognition (HCR). OCR system is most suitable for application like a multi choice examination, printed postal address resolution and HCR is a wider as compare to OCR. This paper gives a detailed overview of different featuresextraction and classification techniques for recognition of sugali script.

2. LITERATURE

At present, the template matching and neural network is popularly used algorithm in the handwritten character recognition. The neural network has a relatively great space to enhance this recognition effect. There are two types of character recognition – Online character recognition and offline character recognition. And offline character recognition is also of two types- optical character recognition and magnetic character recognition. In character recognition the data obtained by scanning is regarded a static representation of a handwriting. The process of identifying the character in a scanned document is called offline character recognition.The offline character recognition because the name itself implies that it involves the automated conclusion of text in a picture into letter codes that square measure usable inside the computer and text processing applications. The use of character recognition and the process of subsequently digitizing the existing information help in easy retrieval, easy access and even easier analysis.

The optical character recognition technique Primarily include 2 steps covering description and classification of a specific character on input pictures, in essence, Associate in Nursing characteristic a picture as explicit character of an alphabet of the language. The characteristics feature of the character that is identified based of which classification or in simpler terms, identification of the character is achieved. The principle aim of this thesis can be described as an endeavour to determine the accurate method for the identification of a sugali handwritten character. The search is basically covering area under neural network and evolutionary algorithms, focusing mainly or identifying the most accurate method that can be used.

The sugali character recognition has received much attention due to variations in the writing style and similarity in the structure of some of the character of the script. The software for character recognition needs adaptability and the ability to solve problems that cannot solve by deterministic method. Search methods and search algorithms is an algorithms for finding an item with specified properties among a various collection of items. For the purpose of problem solving and optimization, the search techniques divided into two categories. The first category is deterministic or classical method and the second category is a non- classical or stochastic search. The non-traditional search approaches are genetic algorithm and simulated annealing.

A typical pattern recognition system consists of three phases namely:

- Data acquisition
- Feature extraction

International Journal of Ethics in Engineering & Management Education
Website: www.ijeee.in (ISSN: 2348-4748, Volume 5, Issue 7, July 2018)

● Classification

Genetic algorithms are basically developed as a computational model which is used for the process of natural evolution. Genetic algorithms are used for search and optimization techniques based on the principles of evolution and natural genetic [1]. In the process of offline character recognitions a two layer feed forward natural network is used as the classifier [2]. The process of character recognition is done by using phases, they are

- Image acquisition
- Feature extraction
- Pre-processing
- Post-processing [3]

OCR is the part of machine Recognition techniques performing an automatic identification [4]. Every Indian script has own composition rules for combining vowels, constants and modifier [5]. Some language which are used for creation of user interface and interfacing with programswritten in other language, like c, c++, java and fortan [6].

3. PROPOSED WORK

In this work, we have proposed extraction techniques,for offline hand written sugalicharacterrecognition. The main focus of research is to simulate the manner in which human beings interact with is other through hand written means of communications. The research is an attempt to use the way in which human recognizes handwritten character and hence utilizes neural network, support vector machine and genetic algorithms. The research work focused especially on accurately differentiates between extremely similar characters of sugali language. The work begins with the collection of sample, i.e. pre-processing and dimensionality reduction and finally completed with character recognition algorithm. The first step of pre-processing included size normalization where the entire image is of a size 32x32 then images where converted to gray scale and background was made uniform.

A large number of features were extracted to represent the input pattern. But the generalization ability of the classifier decreases as the ratio between the dimensions of the data and the number of samples increases. So there is a need for dimensionality reduction to reduce the error probability of the classifierr. The dimensionality reduction process should find a set of dimensions that better captures the variability of the data. The second objective of the work, to model the collected handwritten characters through neural networks and the support vector machines, was achieved successfully. Two types of neural networks were used for the process of classification such as radial basis function networks (RBF) and probabilistic neural networks (PNN). The RBF is a three-layered feed forward network. The number of nodes in the input layer is equal to the number of attributes and the number of nodes in the hidden layer is 100 and the centres of the hidden layer neurons is decided by the nearest neighbour clustering algorithms. The hidden neurons were activated by applying non-linear transformations, the Gaussian kernel. The number of nodes in the output layer is equal to the number of classes and the activation function is linear. The PNN consists of input layer, two hidden layers and an output layer. The number of nodes in the input layer is same as the RBF network. The second layer is the pattern layer, where the output is computed using a probability density function. The comparison of the two classifiers RBF and PNN was also performed in the course of the work. The disadvantages of the two methods RBF and PNN, such as data dimensionality and lower number of samples, can be overcome by the use of another method, the support vector machines the support vector machines are very close cousins of neural networks and the method is a classification and regression prediction tool that uses machine learning Theory to maximize the predictive accuracy while avoiding over fit to data. His work recognition of the character with search approaches is accomplished using Genetic Algorithms (GA). Genetic algorithms, in general, are stochastic search and optimization algorithms which mimic the nature's evolutionary principles to drive its search towards an optimal solution. Out of all the evolutionary principles, the genetic algorithms utilize the most famous principle of all, "the survival of the fittest". One of most striking difference between the classical search and search with genetic algorithms is that the genetic algorithms use a population of solutions in each iteration, instead of a single solution. Since populations of solutions are processed in each iteration, the outcome of the algorithm is also a population of solutions. If the optimization problem has multiple optimal solutions, the GA can be used to capture multiple optimal solutions in its final population. The fitness function was used to calculate the similarity coefficient to each individual. In the work, three different types of fitness functions were used. These different types of fitness functions include.

4. FUTURE SCOPE

This research work in sugali can be used to include the entire character sets of Telugu, for complete word or sentence comprehension etc. It can also be extended to understanding, or rather; identification of handwritten characters of other languages.With the advent of computers into our daily lives, the use of imaging devices such as scanners has considerably increased to allow high speed data entry. Time effective innovations such as optical character recognition devices would not only reduce the work of the user but will also prevent operator wrist problems and other occupational hazards for the users. Optical character recognition techniques might become necessary in the future for interactions with robots. When a robot is faced with comprehension of text, images are taken, which are then converted into text, extracted for their meaning and instruction, and then acted upon.

5. CONCLUSION

Recognition rate is highly affected by similarity of various characters. In this we have treated individual image pixels as features, where each comparisonresults in similarity measure between the input character and the template. The comparison of the character is performed on the by pixel basis. This

proposed is utilizing structural features and decision rules to classify characters. Here we presented work done on handwritten Indian script. First we give all the properties of Indian scripts as well as all the properties of sugali scripts. A lot of work has been done to recognize many languages like Oria, Malayalam, Kannada,Telugu,Tamil,and Gujarati etc.

There is a need for handwriting OCR for Indian script that can help people for communicate the handwritten text like character and numerals to computer processable format. The new method for character recognition will continue appears with the development of a computer technology. In this proposed work there are some techniques which deal with recognition of entire words in state of individual characters.

REFERENCES

[1]. Sourabh S. Patil and Dr. AS Bhalchandra "pattern recognition using genetic algorithm" international conference on I-SMAC (978-1-5090-3243-3/2017)

[2]. SunithaanneM.O.chacko et al, I (IJCSIT) vol 6(3), 2015, 2314-2317 "Offline handwritten character recognitions in south Indian scripts" ISSN:0975-9646

[3]. International Journal of Computer Applications (0975-8887)"a review on offline handwritten recognitions of devnagri script" Volume 117-No. 19, may 2015

[4]. Chaudhri a; mandaviya, k; badelia, p.kghosh, s,2017"optical character recognitions systems" ISBN: 978-3-319-50251-9

[5]. IEEE Transactions on systems, man, and, cybernetics-part c: application and reviews. Vol41, no. 6 november 2011 "offline recognitions of devanagri scripts"R.jayadevan, satsh R. Kolhe.

[6]. Shrutiagarwal,dr. naveenhemarjani, IOSR journal of computer engineering (IOSR-JEC) p-issn: 2278-8727 volume 12, issue 2 "offline handwritten character recognition with devanagri scripts"(may-jume 2013), PP 82-86.