Volume-4, Issue-9, Sept-2015 • ISSN No 2277 - 8160

Research Paper | Computer Science

Recognition of Gurmukhi Script Characters Using Image Segmentation

Satwinder Kaur M.Tech Scholar, Department of Computer Science, Lovely Professional University, Phagwara, Punjab (India)

Tejinder Thind Assistant Professor, Department of Computer Science, Lovely Professional University, Phagwara, Punjab (India)

ABSTRACT

Image processing is a technique to extract the important features of an image and use them in desired manner. It helps us to visualize and analyze image in a better way. One of its main applications is Image segmentation in which we divide an image into desirable parts. In this study, Image segmentation is used for the recognition of Punjabi characters as nowadays there is a need of handwritten documents to be used in digital form. In this proposed work, Back propagation algorithm with Neural Networks is used for the character recognition. As neural network is one of the best methods for the recognition of the handwritten characters. Neural network is trained for the recognition of words that then match the input with the dataset and gives effective results.

KEYWORDS : Thresholding, Image processing, Image Segmentation, Gurmukhi

1. Introduction

Image processing is the process that deals with the extraction of important features of an image from the original image. The main input of this process can be any type of image. It helps us to process an image with its all features. We can also say, it is the process of converting the image into digital form and perform many operations on it so that we can extract the desired information from the input image. Image processing mainly refers to Digital Image processing. Nowadays, it is rapidly growing technology that is used in many areas of business. One of its main applications is image segmentation.

1.1 Image Segmentation

Image Segmentation refers to dividing an image into several parts so that we can refine the image in a better way. To extract important information from the image we first segment it into multiple parts so that extraction can be done efficiently.

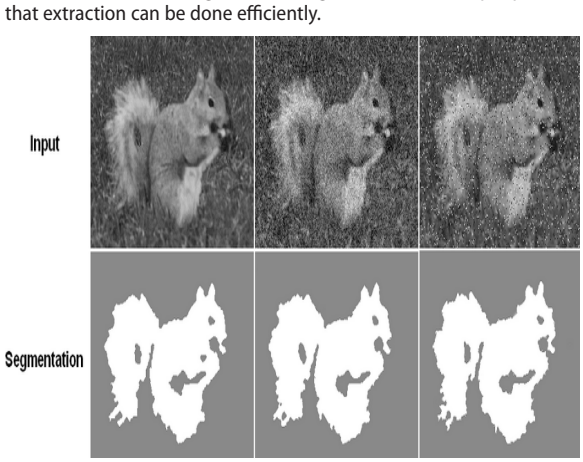

Figure 1: Segmentation of an object

Image segmentation is useful in many areas which select the homogenous parts of an image together to do as desired. There are many algorithms of image segmentation used these days.

1.2 Applications of Image Segmentation

There are different applications in which image segmentation technique is used. These are:

- Medical Imaging
- Object Detection
- Content Based Image Retrieval
- Machine Vision
- Recognition Tasks
- Video surveillance
- Traffic Control Systems

1.3 Methods of Image segmentation

There are various methods of image segmentation from which some of the widely used methods are:

- Thresholding
- Compression based methods
- Edge detection methods
- Region based segmentation
  - Region growing methods
  - Region split and merges method

2. Overview of OCR

OCR (Optical Character Recognition) is the software that helps us to convert the handwritten or printed text into machine readable form. As there are many documents like magazines, books, articles, etc that are stored in the organizations to store historical data but when it needed it is very difficult to find that data as it takes too much time to search the data. So, OCR system is developed to extract the printed texts into digital form. It takes the scanned images of the desired text and easily changes into the digital form.

2.1 Working of OCR with ICR

Humans beings are God gifted with the ability to recognize what we see by our eyes. We can easily recognize every printed or handwritten paper but the computer has no such type of ability because of lack of intelligence. It becomes possible by using OCR system which gives output in digital form of all the scanned images. But it is difficult to OCR too to recognize each and every handwritten document because everyone have their own writing styles even printed documents are also not the same. So to recognize all types of documents, Pattern recognition method is developed which helps to recognize specific patterns.

After this, feature extraction method, also called as ICR (Intelligent Character Recognition) helps to recognize the characters by using its features. Firstly it extracts the features of the handwritten or printed text and then applies pattern recognition method on it to get better results as compared to OCR

2.2 Types of OCR

There are mainly four types of OCR which are used for their own individuals work. OCR mainly works in the offline mode which takes the input as static documents as image. The types of OCR are given below:

1. OCR: It takes the typewritten text or printed text but only one character at a time.

2. OWR: It also takes printed text but it takes one whole word at a time.

3. ICR: It takes the handwritten printed texts which takes one character at a time and involves machine learning.

4. IWR: It also takes the handwritten printed texts which takes one word at a time and involves machine learning. It is especially used for handwritten languages where characters are not separated in the script.

2.3 Techniques used in OCR

There are mainly three types of techniques on which OCR works.

i. Pre-processing: OCR preprocess the input image for enhancing the chances of recognition process.

ii. Character Recognition: It includes mainly two types of OCR algorithms which generate list of characters.

a) Pattern Recognition: in this algorithm, there is a comparison of the input image on the basis of pixels of the character. This algorithm is also known as Pattern Matching or Matrix Matching or Image Correlation.
b) Feature Extraction: this algorithm break down the character into multiple features like lines, line directions, closed loops and line intersections.

iii. Post-Processing: It generates the output as a list of words or characters effectively that are present in the given document which is in the printable or handwritten format.

On similar lines, we are using recognition system in our proposed work by using the image segmentation technique.

3. Objectives of the Study

There are many problems faced by the users when they need to convert the handwritten papers into digital form. For this, they have to type all the characters by their own that consumes lots of time. At that time, there is a need of some recognition systems that recognize all the characters and convert them into digital form.

The main objective of our system is to provide the ease of reading handwritten Punjabi characters by converting them into digital form. It is implemented with the help of back propagation algorithm with neural networks that helps to understand human Punjabi language handwritten characters and recognize them with better results.

In the proposed work, back-propagation algorithm would be used to solve our problem with neural network. BP is one of the best algorithms that would be used with Neural Network techniques.

Artificial neural network is a tool that is used for the computational work. This tool is just like a collection of neurons which are interconnected with each other for responding their results to each other. Artificial neural networks are trained for the particular task which will have to be solved. They have the ability to learn by giving number of desired inputs. They have the neuron interconnections which have some individual weights (in digits) which will vary according to the experience it gets and adaptable for the inputs given to it and capable to learn the desired work.

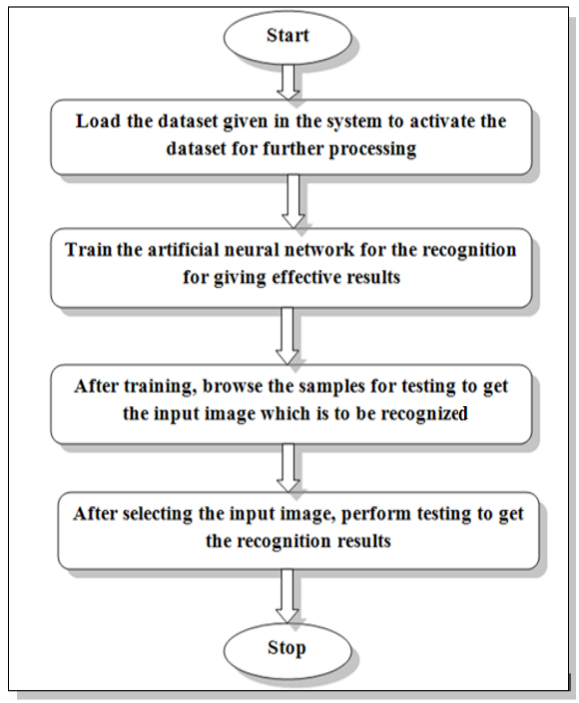

Figure 2: Flowchart of Actual Work

Implementation Steps:

a) Load the dataset created in the system.
b) Then train the artificial neural network to get the desired dataset. In this process, different phases which are present in the actual recognition process are going to be executed. In these phases, feature extraction and character segmentation will be done.
c) Browse the samples for the testing process to get the input image which is going to be recognized.
d) At last, perform the testing process on the input which will give effective results.

At the end, the final result of recognition process is executed with better effectiveness.

This system will be efficient method to recognize the Punjabi characters.

4. Conclusion and Future Scope

There are many character recognition systems present which are used to recognize characters written in different languages. Punjabi characters are used to recognize by the system in this study. This system will understand the difficulties faced by the people during the recognition of Punjabi characters and try to solve their problems by using this system.

In the future work, this system can be enhanced for better performance and better results. This system can also be used for the other languages like English, Devnagari script or some other languages.

REFERENCES

[1] Ashok Kumar, Pradeep Kumar Bhatia, "Offline Handwritten Character Recognition Using Improved Back-Propagation Algorithm." International Journal of Advances in Engineering Sciences Vol.3 (3), July, 2013 | | [2] D.kavitha, P.Shamini, "Handwritten Document into Digitized Text Using Segmentation Algorithm." Special Issue, 4th National Conference on Advanced Computing, Applications & Technologies, May 2014 | | [3] D.M.D.S.S Dassanayake, R.A.D.D Yasara, H.S.R. Fonseka, E.A HeshanSandeepa and L.Seneviratne, "Panhinda - Offline Character Recognition System for Handwritten Articles." 978-1-4799-2845-3/13, 2013 IEEE | | [4] Faisal Mohammad, Jyoti Anarase, Milan Shingote, Pratik Ghanwat, "Optical Character Recognition Implementation Using Pattern Matching." (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2), 2014, 2088-2090. | | [5] Parika Mangla, Harleen kaur, "An End Detection Algorithm for segmentation of Broken and Touching characters in Handwritten Gurumukhi Word." 978-1-4799-6896-1/14, 2014 IEEE | | [6] Shaham Shabani, Yaser Norouzi and Marjan Fariborz, "Handwritten Objects Recognition using regularized Logistic Regression and Feedforward Neural Networks" 2014. | | [7] Tahsina Hashem, Mohammad Asif and Md. Al-Amin Bhuiyan, "Handwritten Bangla Digit Recognition Employing Hybrid Neural Network Approach." 16th Int'l Conf. Computer and Information Technology, 978-1-4799-3497-3/13, 2013 IEEE |