
CONVERSION OF HANDWRITTEN DATA INTO DIGITAL FORMAT

Dharinikaa A¹

¹M.Sc., Department of Computer Science, Fatima College, Madurai, India.

ABSTRACT

In conversion of handwritten data into digital format, the important issue is to make the system recognition of words. Because each individual has various types of handwriting style. This paper presents an approach to recognition of handwritten words and digit based on machine learning technique. The ability to develop an efficient algorithm that can recognize handwritten words and digits. The input is provided by users by the way of a scanner and other digital devices. The paper presents a new application that can turn handwritten page into computerized version by using Intelligent Word Recognition(IWR). It will be very useful to read the ancient script.

Keywords: Intelligent Word Recognition, Pre-processing, Segmentation, Recognition engine

1. INTRODUCTION

This paper refers to the system which convert the handwritten text into the digital version. Handwritten character recognition is a bit difficult job to do, because of different variation of the writing styles. Identification of various style of text using different kind of method is one of the key features in this area of text recognition. It is because of the wide range of writing style. The system must robust to improve the extraction and performance of the system. Nowadays the handwritten text recognition has gained a lot of interests of the industries and its application in various fields. This system uses the Intelligent Word Recognition(IWR) technology which makes the system less complex and makes the process easy to recognize the meaningful word/sentences. Because of this technologies the digitization of the text has become simple. The group of neurons helps capturing the pixel of the input image of the text and extracts a meaning full text/digit by using the Intelligent Word Recognition. Further the data are passed on and transformed into a meaningful manner according to the designed network pattern. Again, the different processes are done to extraction or receive the expected output or to show that character was read at the input side.

2. METHODOLOGY

The process that the system undergoes for the conversion of handwritten data into digital format are Pre-processing, Segmentation, Recognition engine and Post processing

Image Pre-Processing: Pre-processing, which includes techniques such as smoothing, binarization, skew detection, slant correction, thinning or skeletonization, filtering, base line identification, etc., is one of the most important steps in OCR. Pre-processing is a preliminary step that transforms the incoming data into format that can be processed more easily and effectively.

It boosts the effectiveness of recognition systems. Following, the step demands precision because it has a direct impact on the effectiveness and dependability of the other phases. It gets the picture ready, for further operations like segmentation and feature extraction.

Segmentation: In order to determine what exactly makes up the input image, the pre-processed input data must be divided/segmented into sub-data.

This process is called as segmentation. During this segmentation phase, the input material is divided into paragraphs, paragraph is further divided into lines, lines are divided into words, words are divided into characters, and finally characters are divided into sub-characters. The sub-parts are designed so that the subsequent processes may process them more fast and efficiently.

Recognition Engine: Intelligent Word Recognition (IWR) is the recognition of unconstrained handwritten words. IWR engine recognizes entire handwritten words / phrases instead of character-by-character, like its predecessor, optical character recognition (OCR). IWR technology matches handwritten or printed words to a user-defined dictionary, more importantly reducing character errors encountered in typical character-based recognition engines.

Post Processing: In post processing, digitized words are formed after the appropriate analyzation of words. In this process it also checks the semantic meaning of the text. Sematic Analysis is the part of Natural Language Processing.

3. CONCLUSION

The main perspective of this paper is to recognition of handwritten text-lines with the better accuracy. The main focus of the investigation is to improve the system to identify any other character and tried to make our system more font independent that our system can read or recognize any other font. An Intelligent Word Recognition(IWR) and the segmentation algorithm techniques has obtained good results.

4. REFERENCES

- [1] N Arica and F Yarman Vural, “An Overview of Character Recognition Focused on Off-line Handwriting”, IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews
- [2] Verma B K, “Handwritten Hindi Character Recognition Using Multilayer Perceptron and Radial Basis Function Neural Network”, IEEE International Conference on Neural Network
- [3] Jamshed Memon, Maira Sami, Rizwan Ahmed Khan , “Handwritten Optical Recognition (OCR): A Comprehensive Systematic Literature Review (SLR)”. Faculty of IT, Barrett Hodgson University1 January, 2020