Converting Hand-written Content of Scanned Images to Digital Documents

Prof. Gajanan Tikhe ,1. Harshada P. Zade, 2. Yogeshree G. Nikude , 3. Harshal U.Tingne, 4.Apurva S. Pole , 5.Mohini M. Dukare

Department of Computer Science & Engineer, Datta Meghe Institute of Engineering Technology and Research Wardha , Maharashtra .

ABSTRACT

This paper presents an innovative idea of converting different human handwriting for example cursive into Digital text using Conventional Neural Networks, Optical Character Recognition and Intelligent Word Recognition (IWR).we aim to build computer application which will instinctively read the contents from scanned images. Challenges involved in this project was recognizing the various types of characters of human handwriting as well as quality of images . though we have move towards the digitalized world and everyone is talking about digital transformation ,organizations will always have a stack of files or documents as a part of their business. so it is very important nowadays to store the handwritten documents into computer readable form. Thus we need to find perfect character recognition algorithm to perform digitalization of this documents.

Keyword : *text recognition ,converting text***.**

1.Introduction

As our paper title denotes the conversion of handwritten text into legible digital text documents we have used Intelligent Word Recognition to find the text from scanned images by using artificial intelligence. this consist of Conventional Neural Networks .The Neural Network for handwritten Recognition is a group of input neurons may be activated pixels as an input image representing each letter. we have also used Optical Character Recognition (OCR). OCR is a technology that is used to recognize text in images. It is used to convert typed, handwritten or printed documents into machine readable form .Handwritten Text Recognition is very difficult due to great variations of handwritings different size of characters and sharpness of fonts .hand written text recognition is a technique to receive scanned images as input .[1] (Tikhe G. N., 2012)As there are number of historical books ,newspaper which is in paper or print format needs to be store in computer storage disks so that we can use later by searching methods. there is a risk of damaging of these books eventually as a result in loss of important documents.

So to avoid this problem we have come up with the solution .Though we have move towards the digitalized world and everyone is talking about digital transformation ,organization will always have stacks of paper documents as a part of their business. The application of OCR is used to store this files in an organized way .[2] (Tikhe M. G., 2011)A lot of industries need this application for automation .banking is one among them where this application can be used to process handwritten cheques.

Literature Review

Yang et al. has proposed a novel on adaptive binarization based on wavelet. this approach was faster and can be used in real time processing.

Sanakaran et has proposed a novel recognition approach that result is 15 % decrease in word error rate.

2.Methodology

In this we will see in detail explanation of our application .Conversion offhand written text as scanned images into digital text application. receives the scanned images as input and convert it to computer readable form to convert it into digital `

2.2Character Recognition Algorithm

The algorithm used in Character Recognition can be divided into three categories: Image Preprocessing , Conversion, Segmentation ,Feature Extraction And Classification.

1. Image preprocessing is crucial in text recognition for correct character prediction this method typically include noise removal cropping, scaling .First we need to scanned the image from which we wish to convert text to digital text . for this we have used Open CV library .Open CV is a python library that we will be using for preprocessing images .it is a library that is aimed at real time computer missions.
2. considering the problem of character recognition we need to reduce noise as much as possible for this we need to do,[3] (Manwatkar, 2015) Conversion of image to gray scale we need to convert our image to gray scale as sometimes because of low quality of image we would able be to convert only major part of text accurately. But other part of text need to be extracted accurately in a better way. to address this the best way is to alter the image in such a way that the text quality is amplified.
3. In segmentation stage , a sequence of characters is segmented into sub images of individual character .[4] (arora, 2019)each character is resized int 30x20pixels .
4. In Feature extraction the feature of input data are the measurable properties of observations ,which is used to analyze or classify these instances of data.[5] (kartikeyan, 2014) the task of feature extraction is find or search the features in database . and if found return it in output.

5. This stage is decision making stage of the recognition application the classifiers contains the two hidden layers . ,[6] (Manwatkar, 2015) using a log sigmoid function to train the algorithm.

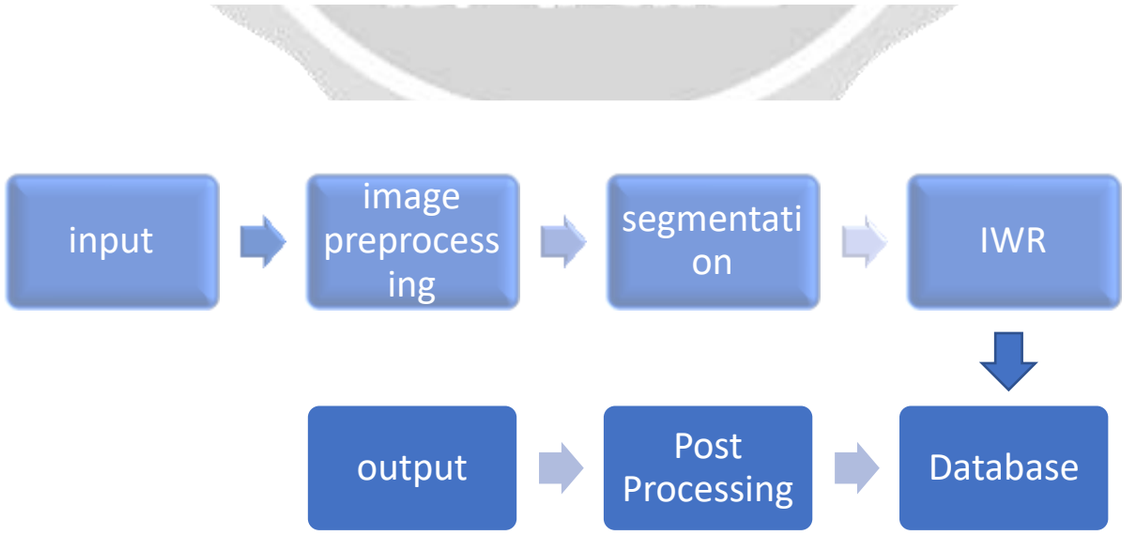

Fig : Conversion of handwritten text as scanned image to digital document module.

3.Conclusion

In this paper we have presented Intelligent Word Recognition (IWR) and segmentation algorithm. and have been used and presented intermediate results. According to our result we have come to the conclusion that more is our module is trained results accuracy increases. We have also provided accuracy rate in our module output. that is each time we are comparing the error rate and returning the less error value. This algorithm is easy to understand not more complex and can detect characters as well as numbers too. It can efficiently remove noise from images as well .more further hope that research on this topic can help in giving maximum accuracy.

4.References

[1] Mayil S. and Vanitha M,"A survey on privacy Preserving data Mining Techniques ",International Journal of Computer Science and Information Technologies. Vol.5 (5), pp. 6054-6056 ISSN: 0975-9646,2014.

[2] G.S Reddy, P. Sharma , S. R .M .Prasanna, C. Mahanta , and L.N. Sharma, "Combined online and offline Assamese handwritten numeral recognizer," in National Conference on Communications ,NCC 2012, 2012.

[3] A. Arora and A.M Namboodri ,"A hybrid model for recognition of online handwriting in Indian scripts ,"in International Conference on Frontiers in Handwriting Recognition , ICFHR 2010,pp.1-5,2013.

[4]U. Karthikkeyan ,M. Vanitha ,"A study on Text Recognition using Image Preprocessing with Datamining Techniques," International Journal of Computer Science and Engineering .Vol-7,Issue-2,Feb 2019.

[5]Pratik Madhukar Manwatkar , Kavita R. Singh ,"A Technical Review on Text Recognition from images ,"IEEE Sponsored 9th International Conference on Intelligent systems and control(ISCO)2015.

[6]Mr. Gajanan N. Tikhe , Mr. Jeetendra Ambulkar , "A Certificate-Based Scheme to Defend against Worm hole Attacks in Ant based Adaptive in Ant based Adaptive Multicast Routing protocol for MANET" , in ICCCT , Delhi on 6th August 2011.

[7]Mr. Gajanan N. Tikhe , Mr.Yogandhar Pandey , "A Secure Scheme to avoid Worm hole attacks in Ant based Adaptive Multicast Routing protocol for MANET" , IFRSA'S INTERNATIONAL JOURNAL OF COMPUTING (IIJC) Volume 2, Issue 1, ISSN(print):2231:2153 , ISSN(online):2230:9039, Jan 2012