

CONVERSION OF IMAGE FILE TO DOCUMENT FILE

S.Christiya¹,S.Jenifer Flora Vinnarasi²,U.Keerthana³,K.Mythili⁴

Department of Computer Science and Engineering

Students^{1,2,3},Assistant Professor⁴

Kingston Engineering College,Vellore,India

ABSTRACT

This Project demonstrates How Image can be converted to portable document format (PDF).These files are self-contained documents that have the same appearance on screen and in print, regardless of what kind of computer or printer are used, and regardless of what software package was originally used for their creation. PDF files are compressed documents, invariably smaller than the original files, hence allowing rapid dissemination and download. This framework has the priority to detect the text in photos and videos. Text recognition is the process of detecting text in images and video streams and recognizing the text contained therein. Once detected, the recognizer then determines the actual text in each block and segments it into lines and words. The Kohonen algorithm that detects text in Latin based languages(French, German, English, etc.) using synthesizer in real-time, on device. Then the modified document is again converted into image file.

KEYWORDS: OCR, Android, Tesseract, Cropping, Editing.

I.INTRODUCTION

OCR also called Optical Character Reader is a system that provides a full alphanumeric recognition of printed or handwritten characters at electronic speed by simply scanning the form. Forms containing characters images can be scanned through scanner and then recognition engine of the OCR system interpret the images and turn images of handwritten or printed characters into ASCII data (machine-readable characters).

Therefore, OCR allows users to quickly automate data capture from forms, eliminate keystrokes to reduce data entry costs and still maintain the high level of accuracy required in forms processing applications.

The technology provides a complete form processing and documents capture solution. Usually, OCR uses a modular architecture that is open, scalable and workflow controlled. It includes forms definition, scanning, image pre-processing, and recognition capabilities. This project has two motivations, a desktop application and Android based mobile application.

Scanned documents, pictures stored in Android devices, and pictures taken by an Android device containing a text are scanned by this OCR. This application will allow its users to retrieve text from these documents to make it editable or to reuse it to perform other actions. For example, people can scan any ID Card, and store its data in their phones directly, or scan books or any important article found in a newspaper and retrieve the needed information and store them in their devices in few minutes.

II. EXISTING SYSTEM

Neural networks are particularly useful for solving problems that cannot be expressed as a series of steps, such as recognition patterns, classifying into groups, series prediction and data mining. The neural network is presented a pattern. This could be an any type of image, a sound, or any other sort of data. OCR technology is allows the conversion of image which is scanned of printed character into text or any other information that user want using android mobile. OCR technology uses three phases first is Scanning of documents as optical images. Next is Recognition which involves converting those images to character streams representing letters of recognized words and the final element used to accessing or storing the text which are already converted. The user begins by capturing an image using mobile camera containing text .Most of the character recognition systems will be recognized through the input image with computer software.

III. LIMITATIONS

1. Large amount space require for computer software and scanner.2. Converted text is nothing but only the extracted text and cannot be modified.

IV. LITERATURE SURVEY

1.Ankur Gupta_ and Lyle N. Long , “ Character Recognition using Spiking Neural Networks “ Presented at IEEE Neural Network Conference, Orlando, FL, Aug.,2007.

Our eyes are recognition the luminous pattern of printed character and our brain is using this to figure out what we are trying to say. Apart from humans nowadays even the computer or capable of performance this task using the technique called OCR. OCR helps in bringing the text available in analog format to digital format.

2. Recognition of Handwritten Devnagari Characters through Segmentation and Artificial Neural networks International Journal of Engineering Research & Technology vol. 1 Issue 6, August – 2012 ISSN: 2278-0181.

The objective of the paper is to utilize this feature of the computer through an android app. This visual capability is brought out using a android mobile phone working on tesseract OCRANGE the android app provides user to recognition the text from either an image stored in the gallery, image taken with a camera, from a stored document mobile or allows to store a name of the locations from the map application available in mobiles,

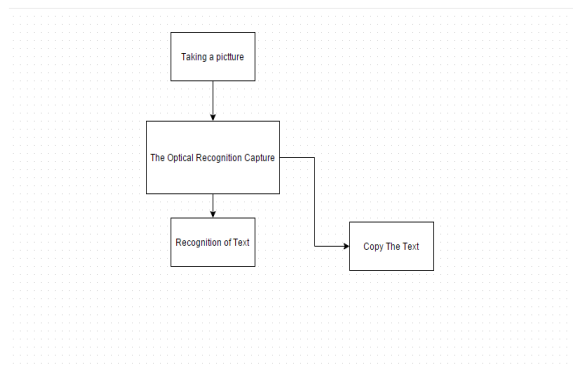
3.Journal of Pattern Recognition Research 2 (2011) 269-277 Received December 12,2009.Accepted September 14,2011.

This paper tells about OCR system for offline hand written character recognition. Preprocessing techniques used in document images as an initial step in character recognition system based presented. The feature extraction step of optical character recognition is the most important. It can be used with existing OCR methods, especially for English text.

V. PROPOSED SYSTEM

In this proposed System the OCR takes image as the input and get text from that image. The character recognition method is presented by using OCR technology and higher quality camera of android phone. OCR technology is used for pattern recognition of characters. Neural network will recognize the complete character with the help of Kohonen algorithm. By using this algorithm the images can be scanned and the text can be extracted from the PDF files then the data is modified. The Data is modified using text editor API which can be implemented using GitHub. GitHub is a development platform inspired by the way you work. From open source to business, you can host and review code, manage projects, and build software alongside millions of other developers.

VIII. FLOWCHART



IX. CONCLUSION & FUTURE ENHANCEMENT

In this system we have proposed an artificial neural network based simple colour and size invariant character recognition system to recognize english alpha numeric character. Our proposed system gives excellent result for the character letters when they are trained and tested separately but produced satisfactory result when they are processed together. In addition our system is computationally inexpensive and easier to implement and the is that the system will produce the conversion of English text to Marathi text.

REFERENCES

1. Ankur Gupta_ and Lyle N. Long , “ Character Recognition using Spiking Neural Networks “ Presented at IEEE Neural Network Conference, Orlando, FL, Aug., 2007.
2. Recognition of Handwritten Devnagari Characters through Segmentation and Artificial Neural networks International Journal of Engineering Research & Technology vol. 1 Issue 6, August – 2012 ISSN: 2278-0181.

3. Journal of Pattern Recognition Research 2 (2011) 269-277 Received December 12, 2009. Accepted September 14, 2011.

4. Deepali Thombre, Toran Verma, Neural Network approach to recognition online Handwriting script, 1, 2RCET, Bhilai C.G., India IJECT Vol.2, Issue \$, Oct., - Dec 2011.

5. Review of Handwritten Pattern Recognition of Digits and special character using Feed Forward Neural Network and Izhikevich 2014 International Conferences on Electronic Systems.

6. Neila Mezghani, Amarmitiche, Mohammad cheiret, online recognition of handwritten Arabic characters using A Kohonen neural network:

7. Ravina Mithe, Supriya Indalkar, Nilam Divekar, "Optical character Recognition", International journal of Recent Technology and Engineering: