

An Efficient Scheme for Invariant Optical Character Recognition Using Edge Detection

Devanand Kale¹, Sameer Shaikh², Nitin Valake³, Prof. Pankaj Agarkar⁴

Department of Computer Engineering, Dr. D Y Patil School of Engineering Pune, India

Abstract: In this paper optical character recognition is used for implementation of an efficient translation and scale invariant optical character recognition with rotation is presented. which is based on transformation of the image triple-correlation domain and appropriate clustering which is one to one related to the class of all shifted-rotated-scaled versions of the input image and also robust to a variety of noises. Special attention is given to binary images and further used for Optical Character Recognition and simulation results shows the performance of the proposed implementation.

Keywords: Optical character recognition, Image representation, clustering, Transformation.

I. INTRODUCTION

In the running world, there is growing demand for the software systems to recognize characters in computer system when information is scanned through paper documents as we know that we have number of newspapers and books which are in printed format related to different subjects. Now a days there is a huge demand in “storing the information available in these paper documents in to a computer storage disk and then later reusing this information by search process”. Storing documents as image and perform OCR on those images as needed. But the most important difficulty while perform OCR is font characteristics in documents. As a result, while reading the character, computer is unable to recognize it. Document processing is the activity of operating on information captured in some form of persistent medium. Usually, that medium is paper writing, and documents are bundles of paper with information captured in print or in writing. Sometimes in this document processing we need to process the information that is related to languages other than the English in the world. CHARACTER RECOGNITION SYSTEM is used for processing such document

Thus our need is to develop character recognition software system to perform Document Image Analysis which transforms documents in paper format to electronic format. For this process there are various techniques in the world. Among all existing techniques we have chosen Optical Character Recognition as main fundamental technique to recognize characters. IN Optical character recognition conversion of paper documents in to electronic format is an on-going task in many of the organizations particularly in Research and Development (R&D) area, in government institutions and also in large business enterprises, so on. From our problem statement we can introduce the necessity of Optical Character Recognition in mobile electronic devices like mobiles, cameras to acquire images, scanners and recognize them as a part of face recognition

Optical Character Recognition, is a technology that enables you to convert different types of documents, such as scanned paper documents or images captured by a digital camera into editable and searchable data imagine you’ve got a paper document - for example, scanned books, brochure, or any document file contract your partner sent to you by email, or other communication platform. Obviously, a scanner is not enough to make this information available for processing operations such as editing, say in Microsoft Word. All a scanner can do is create an image or a snapshot of the document that is nothing more than a collection of black and white or color dots, which is also known as a raster image. In order to extract and repurpose data from scanned documents, captured images from digital cameras or image-only PDFs, you need an OCR system that would recognize a letter and put them into words, then the words into sentence, that enable you to access and edit the content of original document.[5]

II. PROPOSED WORK

While processing OCR on image need to be converted in grey scale Gray level images contain an enormous amount of data, which is much irrelevant. The general edge detection involves three steps: filtering, detection and differentiation. In the first stage, the image is passed through a filter in order to remove the noise. The differentiation stage highlights the locations in the image where intensity changes are significant. In the detection stage, those points where the intensity changes are significant are localized. Edges characterize object boundaries and are useful for segmentation (process of partitioning digital image in to segments) identifies objects in a scene. Edges of an image can be detected using a periodical convolution of the function f with specific types of matrices m .

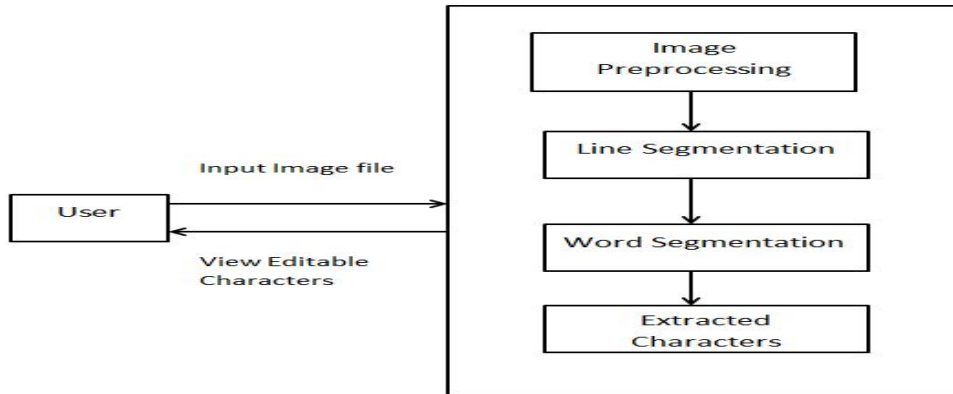


Fig 1. System Architecture

There are four model:

1. Image preprocessing
2. Line segmentation
3. Word Segmentation
4. Extracted Characters

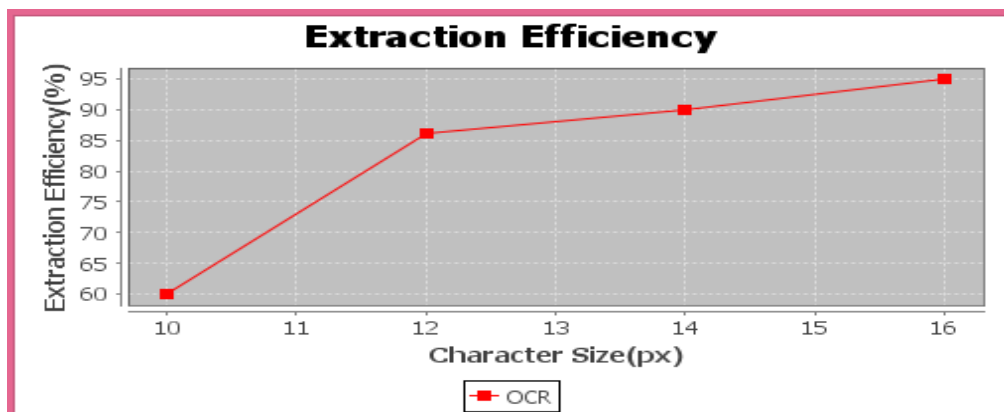
Image pre-processing: The input image file is taken and all the noise (blurred image, unwanted pixel) present in it are removed.

Line segmentation: Edges of the characters are recognize by creating a lines over it.

Word Segmentation: Word segmentation use the TRIPLE CORRELATIONS algorithm.

Extracted Characters: Hence at the end we get the required character in the soft copy (word, txt) format.

III. RESULTS



Graph

In Graph, efficiency of extraction is proposed as per the character size in pixel. When character size is 10px then extraction efficiency is 60%, when character size is 12px then extraction efficiency is 86%, when character size is 14px then extraction efficiency is 95%, result suggest that efficiency of character extraction is dependent on font size.

IV. CONCLUSION

In this paper, a simple and efficient OCR system for Basic English characters is presented. Proposed methodology of system is capable of handling different font sets and sizes. Character set is all English so the methodology developed can be extended to any other sub- English language like Spanish or Mexican with further developments in the system design.

REFERENCES

- [1] MAJIDA ALI ABED HAMID ALI ABED ALASADI Simplifying Handwritten Characters Recognition Using a Particle Swarm Optimization Approach EUROPEAN ACADEMIC RESEARCH, VOL. I, ISSUE 5/ AUGUST 2013 ISSN 2286-4822
- [2] Chirag I Patel, Ripal Patel, Palak Patel Handwritten Character Recognition using Neural Network International Journal of Scientific & Engineering Research Volume 2, Issue 5, May-2011
- [3] Dileep Kumar Patel, Tanmoy Soml, Sushil Kumar Yadav, Manoj Kumar Singh, " Handwritten Character Recognition Using Multiresolution Technique and Euclidean Distance Metric" JSIP 2012, 208-214
- [4] Amir Bahador Bayat Recognition of Handwritten Digits Using Optimized Adaptive Neuro-Fuzzy Inference Systems and Effective Features Journal of Pattern Recognition and Intelligent Systems Aug. 2013, Vol. 1
- [5] http://www.ocrtoword.com/a01_ocr_technology_free_ocr_to_word.html.
- [6] Automatic Number Plate Recognition S .Kranthi, K. Pranathi, A. Srisaila Information Technology, VR Siddhartha Engineering College, Vijayawada, India kranthi_devara@yahoo.co.in, pranathi.pamarthi@gmail.com, sr.saila@gmail.com
- [7] Optical Character Recognition using Neural Networks Deepayan Sarkar University of Wisconsin{Madison ECE539 Project, Fall 2003.
- [8] International Journal of Recent Technology and Engineering (IJRTE)ISSN: 2277-3878,Volume-2, Issue-1, March 2013.