

Proceedings of National Conference on Innovative Trends in Computer Science Engineering (ITCSE-2015) held at BRCMCET, Bahal on 4th April 2015

Extracting Text Part Using MATLAB

Poonam Rani, Payal Taneja Daulat Sihag

Deptt. Of Computer Engineering Jan Nayak Choudhary Devi Lal Memorial College of Engineering , Sirsa Erpoonam78@gmail.com

ABSTRACT --The subject of 2-D and higher dimensional object recognition finds widespread applications in areas such as image registration and pattern recognition. Radon transform is one technique used for efficient object matching

However, so far as we know, no results have been obtained that solves the recognition problem completely in the projection domain due to coupling of transform parameters. We develop a novel method for such parameter decoupling and an improved phase correlation method for accurate practical shift estimation, resulting in a fast matching algorithm based on projection data only. Simulation results show that the proposed algorithm is much

faster than similar state-of-the-art approaches such as that in with comparable estimation accuracy.

Keywords—Fast algorithm, object recognition, pattern recognition, phase correlation method, text extract.

I. INTRODUCTION

Machine replication of human functions, like reading, is an ancient dream. However, over the last five decades, machine reading has grown from a dream to reality. Optical character recognition has become one of the most successful applications of technology in the field of pattern recognition and artificial intelligence. Many commercial systems for performing OCR exist for a variety of applications, although the machines are still not able to compete with human reading capabilities. Optical Character Recognition deals with the problem of recognizing optically processed characters. Optical recognition is performed off-line after the writing or printing has been completed, as opposed to on-line recognition where the computer recognizes the characters as they are drawn. Both hand printed and printed characters may be recognized, but the performance is directly dependent upon the quality of the input documents. Progress in OCR has been steady if not spectacular since its commercial introduction at the Reader's Digest in the mid-fifties.. Our purpose is to examine in some detail examples of the errors committed by current OCR systems and to speculate about their cause and possible remedy.

Overview of our text recognition approach

Given a document image, there are three major steps in our approach for text recognition. First, we extract the text pixels from the input document. For an input image, the user provides example text areas where each text area is a rectangle that contains a horizontal string. The user can rotate the rectangle to select a text string that is not hori- zontally placed in the image. Since each rectangle contains a horizontal string, we exploit the fact that the text pixels are horizontally near each other to identify the colors that represent text in the image and use the identified colors to extract the text pixels. Second, we dynamically group the extracted text pixels into text strings, which is the main focus

of my topic. Third, with the identified text we detect the orientation of each string and rotate the stings to the horizontal direction for text recognition using a commercial OCR product. This topic will focuses on the second step of string identification, which we will describe.

II.WHY MATLAB?

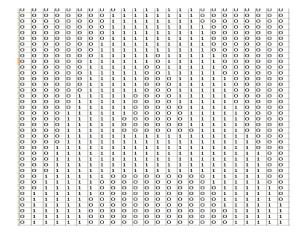
MATLAB stands for MATrixLABoratory. Here you play around with matrices. Hence, an image (or any other data like sound, etc.) can be converted to a matrix and then various operations can be performed on it to get the desired results and values. Image processing is quite a vast field to deal with. We can identify colors, intensity, edges, texture or pattern in an image. In this tutorial, we would be restricting ourselves to detecting colours (using RGB values) only. Using MATLAB you can solve technical computing problems faster than with traditional programming language, such as C, C++, JAVA, FORTRAN. There is a wide range of applications, including signal and image processing, image accusation, Neural Network, etc

Feature Extraction

The techniques for extraction of such features are often divided into three main groups, where the features are found from:

- The distribution of points.
- Transformations and series expansions.
- Structural analysis.

In MATLAB mat2cell command is used for the extraction of image in form of a cell for correlating with the saved templates. Fig. shows extraction of character in Matrix form.



III. RELATEDWORK
Conversion of Grey Image into Binary Image

IJRRA ISSN: 2349-7688

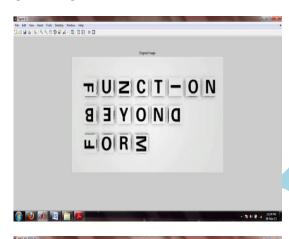


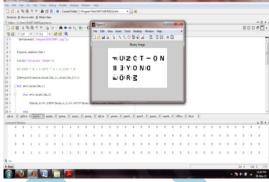


Proceedings of National Conference on Innovative Trends in Computer Science Engineering (ITCSE-2015) held at BRCMCET, Bahal on 4th April 2015

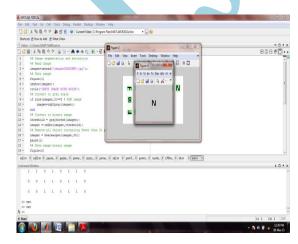
A new binary image is formed by morphological operation on a binary image in which the pixel has a non zero value at Extraction of text part from vehicle image that location in the input image only if the test is successful. A structuring element is a small matrix of pixels with a value of zero and one. An image with structuring element of structuring elements is specified by matrix dimensions. The shape of the structuring element is specified with the help of the patter of ones and zeros.

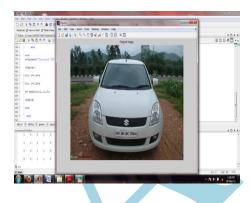
Original Images





Extraction of text part from image









IV. FUTURE SCOPE

New methods for character recognition are still expected to appear, as the computer technology develops and decreasing computational restrictions open up for new approaches. There might for instance be a potential in performing character recognition directly on grey level images. However, the greatest potential seems to lie within the exploitation of existing methods, by mixing methodologies and making more use of context. Integration of segmentation and contextual analysis can improve recognition of joined and split characters. Also, higher level contextual analysis which

IJRRA ISSN: 2349-7688 167





Proceedings of National Conference on Innovative Trends in Computer Science Engineering (ITCSE-2015) held at BRCMCET, Bahal on 4th April 2015

look at the semantics of entire sentences may be useful. Generally there is a potential in using context to a larger extent than what is done today. In addition, combinations of

multiple independent feature sets and classifiers, where the weakness of one method is compensated by the strength of

REFRENCES

- [1] Kumar Parasuraman, Member, IEEE and P.Vasantha Kumar "An Efficient Method for Indian Vehicle License Plate Extraction and Character Segmentation "
- [2] Rachana Chahar , Dr.Jitendra Kumawat & Himanshu G. Bhavsar " Automatic License Plate Recoganization System Based on Image Processing "Volume 4, Issue 4, April 2014
 [3] Thomas Kollar and Jonathan Schmid Optical Character Segmentation
- and Recognition from a Rochester Flag on 05/07/02
- [4] Yannis Avrithis, Anastassios Delopoulos and Stefanos Kollias AN EFFICIENT SCHEME FOR INVARIANT OPTICAL CHARACTER RECOGNITION USING TRIPLE CORRELATION

IJRRA ISSN: 2349-7688 168