

Image Processing and Text Extraction Python

Deepika S

MCA Scholar, School of CS & IT, Dept. of MCA, Jain (Deemed-to-be) University, Bangalore

ABSTRACT

Image processing system is a part of processing the text information and it is embedded or written in image of different forms, the image text process it describing and transferring information which helps the peoples in communication solving problems and availability. This requires image system to be trained as per the human language, so that it can satisfy the needs of the user. When it comes to processing the more human kinds of information like printed book or letter it will increase the complexity of the work which a computer must handle and that is where the OCR comes in the process of image where we can apply the various pre-processing techniques and the algorithm such as Tesseract to recognize the characters and words given image or text document and the text extraction is useful to retrieving the information and editing, searching. These techniques are applied to extract the text accuracy information. We can extract the given text by image both the characters and words in different languages.

Keywords- OpenCV-python, image processing, text extraction, virtual image.

1. INTRODUCTION

Text recognition is a technique that recognizes the text document and data in the desired format. The text recognition process involves the pre-processing and feature extraction and segmentation and also post-processing. The pre-processing is performed as a binarized the image to convert into gray scale. The feature extraction is the text images into identifying characters, words and digits this document describes techniques for converting the textual content of a machine-readable format it analyzes and compare the technical challenges methods and performance of text detection and recognition of the given image or color image. The prior is a classified the step-by step to integrated the sub-problem and the purpose of this paper is to review algorithm and discuss the including text localization and identification of text in both the characters and words for future research.

- It is useful for describing the data or content of given image.
- It can be easily extracted into the text from a given image both the characters and words.
- It is very useful to the human beings to access the data into the image and it consumes time also.

2. LITERATURE REVIEW

[1] In this paper overview of the various text recognition and techniques, the various methods recognition and algorithm has been presented. In this review the various text and recognizing the algorithms accuracy are discussed. According to survey they are providing detailed steps and flow of the text recognition techniques are explained. This paper provides the good survey of recognition and characters in many fields who have begun to work. [2] In this paper explains about the large number of algorithms are been proposed in the literature and we can use the single method to satisfactory to all the performance of applications due to the large variations of characters, font, size and texture. In this paper is deriving the satisfactory results by enhancing the input by the image for the optimum levels of accuracy from Tesseract. [3] In this paper the extraction of the text from the given image or document is important in different areas. In this author has proposed the algorithm it gives the good performance and accuracy, text extraction. They are using OCR (optical character recognition) to extract the text from given image and providing the better accuracy and performance.

[4] In this paper it provides the comprehensive survey of the text information and extraction or extract the text from given image. Even large number of algorithms are proposed in the literature and no single method can provide to satisfactory performance to the all applications due to the large variations in characters font, size, texture. And they are several information sources for text extraction in images, shape and texture. There is lack of public test set makes it difficult to compare the performance of algorithms. [5] In this paper the author proposed the algorithm for solving the problem of offline character recognition. We have to given the input images. The algorithm is trained on the data that was initially present in database. And it has done the pre-processing and detect the line. It provides brief survey of the applications in the various fields. This paper providing good literature and work in the field of OCR (optical character recognition). [6] In this paper provides the overview of the text extraction techniques and algorithm is to be planned earlier. And the proposed system is capable of detecting and recognizing texts from given or various images. This paper explains the additional exposed a performance and compare table and various techniques from the text extraction from an image. It provides more advantage and restrictions. The purpose of this paper is to classify and review the various recent papers and performance analysis for the future analysis. It developing the future work and quick text extraction from an image.

3. PROBLEM STATEMENT

The problem of text extraction TIE system receives the input from given image gray scale or color image and evaluate the character from given image which is text written image and print on document or word file. They are used to indicate extraction of text process, text detection and localization. The terminology defined in our previous publication to maintains consistency. A system that can extract text from many different languages text extraction system to the further, which is composed of the following stages.

4. FLOW CHART

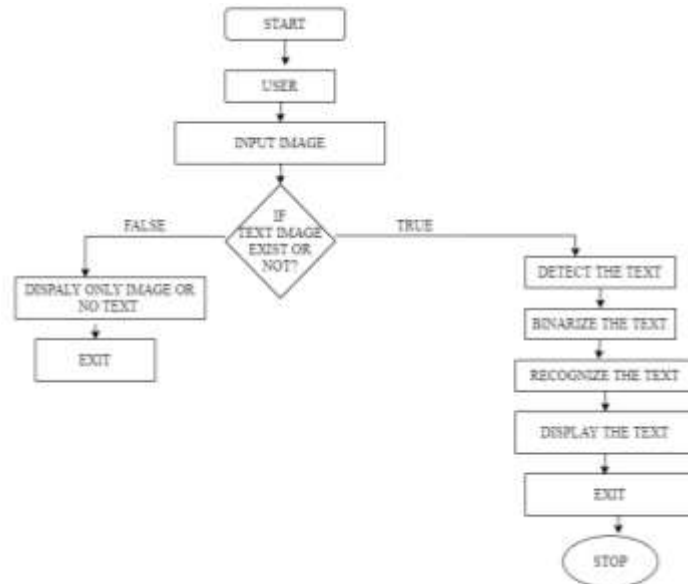


Fig1. Work Image processing and text extraction

In this flowchart we can see the it starts by user and user input the image from JPG and PNG format. We can input image like both ways like text existing image or normal image and the loop consists two ways.

- 1. we can input without text existing image.
- 2. we can input text existing image.

Without text existing image and it identifies image and it extract the image, if no text occurs it displays only image and exit the loop. It input the text existing image and it analyzes the given image and identifies the given image and extract and binarize the given image like displaying the both characters and words and exit and stops the loop.

5. DESIGN

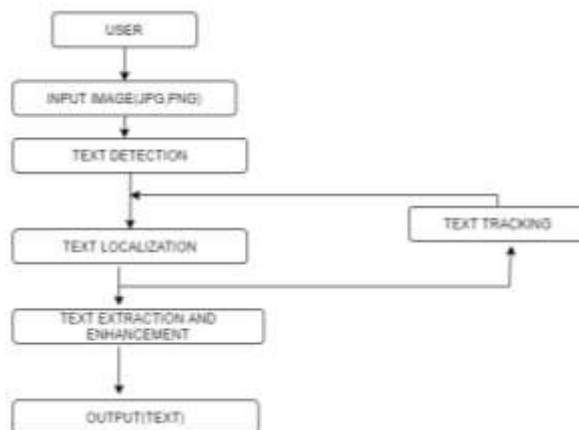


Fig2. Image processing and text extraction design

An image processing text extraction it is one of the pre-processing technique to extract the text from the given image and it is easily recognize the text from the given image using OCR. It provides the user to understand the given information easily. It providing the user to read the both the characters and words. The design of image processing includes as follows.

6. ARCHITECTURE

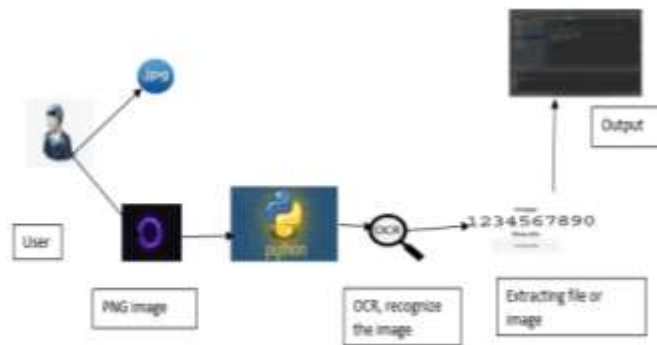


Fig 3. System Module of image processing and Text extraction

7. RESULT AND CONCLUSION

The main objective of the project is to develop an algorithm that will be used to identify the image and it extracting the data or text using OCR and it analyze the text from the given image and it display the like both characters and words.

This project explains about the implemented of the image processing and text extraction using of an optical character recognition to identifies the PNG and JPG image and recognize and extracting the text from given image and it displaying the data from according their spaces and digit number spaces. It provides given image text to both words and characters and we can also implement given image in different languages.

8. REFERENCES

- [1] H.sawney, "Query by image and video content," *IEEE computer*.
- [2] Z. a. S.Y.Tan, "Automatic parsing of news video," *IEEE conference on multimedia computing and systems*.
- [3] K. Prakash, "Text extraction from image using python," *International Journal of trend in scientific Research and development*, 2017.
- [4] M. a. S.M.Stevens, "Intelligent access to digital video," *IEEE Infor media project*.
- [5] M. Y. a. N.Ahuja, "Detecting faces in images," *IEEE pattern analysis and machine intelligence*.
- [6] X. H.Jiang, "Integrating visual, audio and text analysis for news video," *IEEE International conference on image processing*.
- [7] E. a. M. T.Sato, "video OCR for digital news Archive," *IEEE workshop on content based access of Image and video Database*.