International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 07 | July - 2021 ISSN: 2582-3930

Document Scanner Application

Mrs. Asma Begum [1], R Harish[2], Syed Sarosh Shafiq Ahmed[3]

Assistant Professor, Department of Information Science & Engineering, Atria Institute of Technology, Bangalore, India[1]
Student, Department of Information Science & Engineering, Atria Institute of Technology, Bangalore, India[2, 3]

Abstract - With the increase in digital use of applications, the requirement of documents at various checkpoints has become necessary. This paper aims to digitally store all the important required documents into the user's embedded system. This paper includes digital unwrapping of text images for accuracy document scanning , text recognition and edge detection. Documents in today's world stands as a proof for verification of an individual. But a person cannot carry all these documents. The scanning of document and pages in the poor luminance can cause long dark regions alongside of the margin. Imperfect thresholding during binarization can also causes noise blocks. An efficient algorithm is been proposed for effective edge detection of the document and developing application for scanning .

***Key Words*:** Thresholding, binariztion, documents

1.INTRODUCTION

Conversion of a paper document into an image is done either by scanner or cameras. Due to non-uniform illumination while scanning of documents or placing a document scantly in a scanner causes heavy darkness alongside the margin. . Capturing images by digital cameras in poor illumination also produces this type of border noises. The appearance of margin noise might distort the text portion adjacent to the margin and thus segmentation of the region of interest are a difficult task . This marginal noise is composed of both textual noise and non textual noise and it resides along the border of the document. This increases difficulty for edge detection of an document. In this paper an effective algorithm is been discussed to detect edges.

2. METHODOLOGY

There are various algorithms outside which is used to detect edges of an document and to remove the marginal noise from the document. We want to define effective approach for removal and edge detection in the following way:

1. Apply Gaussian blur to filter out some noise.
2. Then there is an edge detection algorithm which detects all the edges in the image.
3. Loop through all the perimeter edges and notice all the closed curves.
4. Find the one with largest area or perimeter.
5. Find the corner (four) points of the area.

OpenCv captures the image in the contour-realize the function of using the mask to cut out the image in PS (Using Opencv extract area circled by contour).I often feel that if photoshop is open source, it will benefit many coders who are engaged in computer vision~ Since it is not open source, let me explore it myself. I hope that from this blog post, I can put it in PS step by step. The function is broken down, welcome to correct and exchange. It is easy for us to use the

findContours() function to extract the contours in the image, but there is no function to output the image surrounded by the contours. Here are a few functions with similar functions:

cvimageroi(): Get the rectangular area surrounding the ROI (Region of Interest), you can directly feed in the contour point set as a parameter and then return the best rectangular point set surrounding this contour, and then use the Mat member function Mat(Rec) to This area is extracted. Unfortunately, only rectangular regions can be extracted in this way, but irregular regions cannot be extracted.

boundingRect() : This is the same function as the previous function, and it also gets a rectangular area.
First, the original image is grayed and binarized, and then the contour is extracted. After the elimination and de-noising process, the contour is full of white by the drawContours() function. The specific approach is :

Mat hole (res.size(), CV_8U, Scalar(0)); layer
cv::drawContours(hole,contours1, -1, Scalar(255), CV_FILLED);
Then use the copyTo() function to overlay the original image on the mask layer, and you're done.

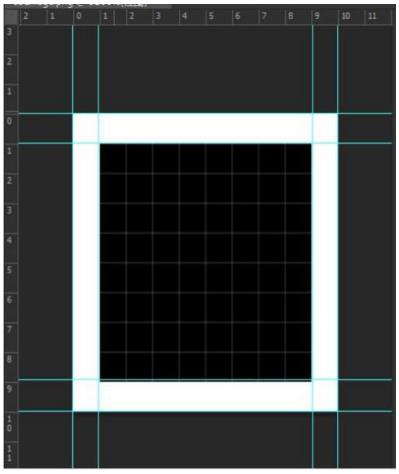

Fig -1: Edge detection

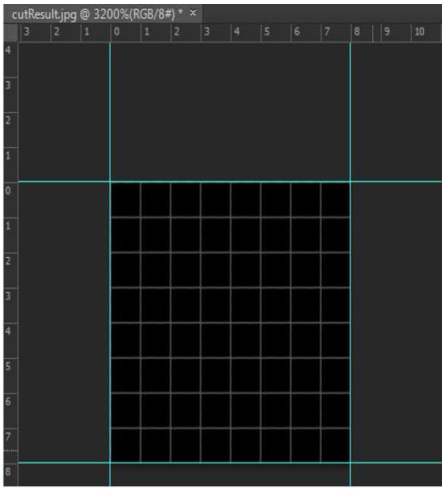

Fig -2:Crop result

3. ARCHITECTURE DIAGRAM

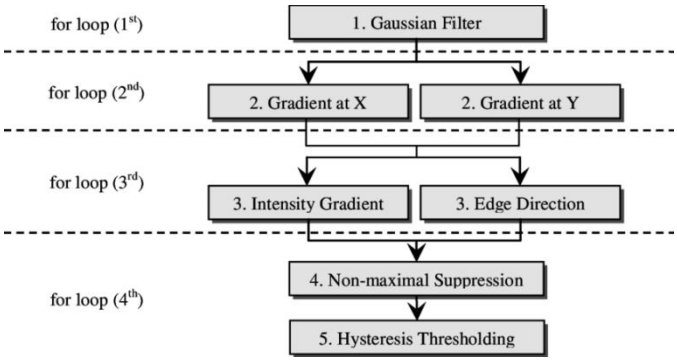

Fig-3: Canny edge detection block diagram

The fIg-3 shows the work flow process of algorithm called canny edge detection.

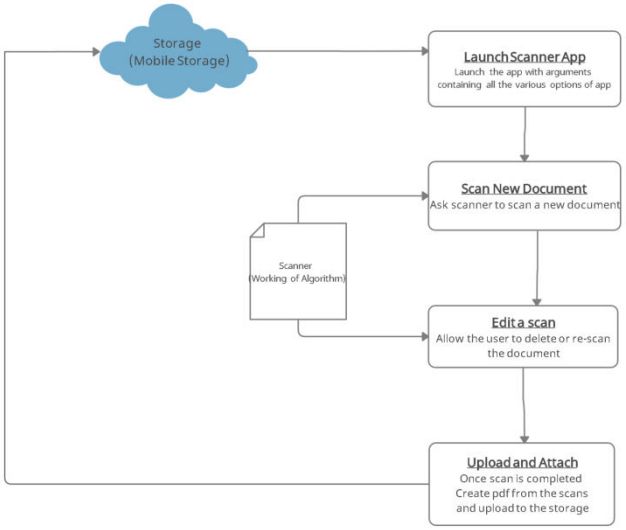

Fig-4: Document scanner application architecture

In the above architecture (figure 4), the architecture shows how the application should be developed and to be used by the end user using it . The user launches the scanner application where the user will be redirected to the camera or gallery to select the pictures to be scanned. The user scans the new document and then the user is asked for either editing the scanned document or go for further scans. The scanned pages are combined to form a portable document format (PDF). the PDF is stored in the local mobile storage for further usage by the end user.

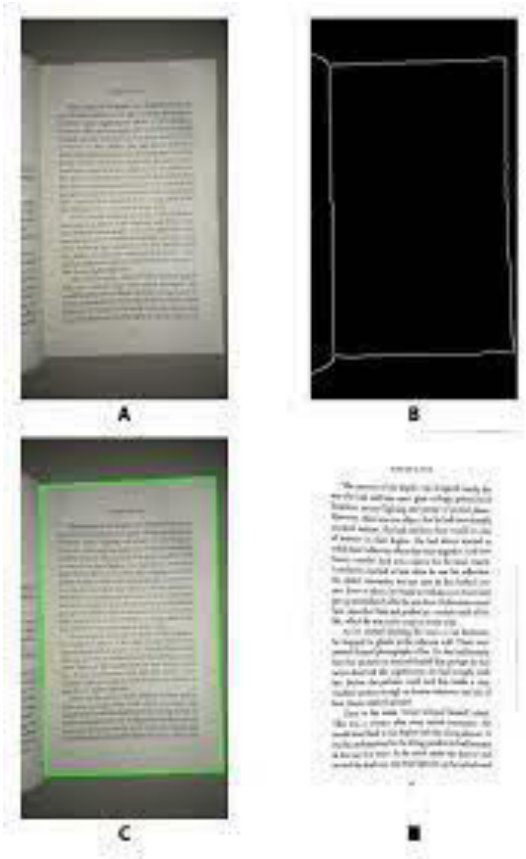

Fig-5: Scanned document using Canny edge detection.

The fig-5 shows successful edge detection of document using canny edge detection algorithm.

4. CONCLUSIONS

An approach as been made to detect edges of an document which is to be scanned by reducing marginal noises caused by poor intensity and low quality pictures. In this paper an efficient algorithm as been used to detect edges of an document which is canny edge detection algorithm. And also an overview is been given on how develop and implement an document scanner application using canny edge detection.

REFERENCES

1. Arpita Dutta, Arpan Garaiand Samit Biswas (2018)-"Segmentation of Meaningful Text-Regions from Camera Captured Document Images".

2. Shaodi You, Yasuyuki Matsushita, Sudipta Sinha, Yusuke Bou, Katsushi Ikeuchi (2017) - "Rectification of folded documents".

3. Sabila Nawshin, Samir Kumar Das, Najia Hossain, Joyita Das Mela, Abu Shafin Mohammad Mahdee Jameel, Salekul Islam-(2019) - "Development of an automated lowcost book scanner and translator".

4 . Quang Nhat Vo, So Hyung Kim, Hyung Leong Yang , Guee Sang Lee (2018) - "Text segmentation using fully convolutional network in handwritten document images".

5. Kushagra Mahajan, Monika Sharma, Lovekesh (2019) - "Character keypoint based homography extraction in scanned documents for efficient information extraction".