International Journal of Research in Engineering, Science and Management
Volume-1, Issue-12, December-2018
www.ijresm.com | ISSN (Online): 2581-5792

Design of Text Recognition in Scene and its Translation into Different Languages

Ghodke Manisha[1], Sonawane Harshada[2], Deshmukh Dattatray[3]
[1,2,3]*Student, Department of Computer Engineering, D. Y. Patil School of Engineering Academy, Pune, India*

***Abstract*: In this project we proposed system which used mobile camera for capturing text and it convert that text into user defined language as well as audio format. Recognition of scene text is a challenging problem compared to the recognition of document which is printed. a approach is proposed to recognize text in complex background natural scene, formation of words from text which is recognized, checking spelling and translating words into user defined language and finally overlay translated word onto the image. The field of image processing has increased up to a new dimension because of availability of high resolution cameras. Now technology is rapidly growing, various technologies are being developed for handheld devices like tablets and Smartphone. Recent years, huge popularity gained by Android platform in terms of number of available applications and market share. Propose system help to blind persons read text and packaging of product from hand-held devices in their daily lives explore user interface issues and assess algorithm robustness in extracting then reading text from different objects with complex backgrounds.**

***Keywords*: translatione, extraction, recognition, robustness.**

1. Introduction

OCR is Optical character recognition. The pre-processing stage sets the accuracy level for the forthcoming stages, but the available state-of-the-art pre-processing algorithms have the drawback of high run times. Thus, the appearance of preprocessing block to suggest an 'accuracy speed' mutual exclusion. A standard pre-processing stage involves: layout analysis and binarization, The process of character recognition of any script can be broadly broken down into three steps-classification, post-processing,t ext extraction. Typical text extraction includes a collection of operations that apply efficient transformations on an image. It takes in a raw image, removes noise, distortion, skewness and detect text segment by applying various segmentation & connected component analysis and feature extraction techniques.

The stage of classification is the main decision making stage of a text recognition system and uses the features extracted in previous stage to identify the text segment according to preset rules. The post-processing stage, which is the End stage, the process of refining the decisions taken by the previous stage and recognizes words by using context provide result as improved recognition. It is totally responsible for outputting the best solution and is often implemented as a collection of techniques that rely on character frequencies, lexicons, and other context information Scene images are characterized by complex background, perspective distortion, low resolution and poor quality. Scene text can appear with any slant, tilt, in any light and upon any surface and hence hard to detect, localize and recognize. Number of pre-processing steps performed by the actual recognition in the implementation of text recognition system. In this paper, a novel approach is proposed that allows recognition of texts from color images with complex backgrounds and translate recognized word into user defined language and overlay it onto the image. With respect to different kinds of text appearances, including font style, font color font size etc. approach is targeted towards being robust.

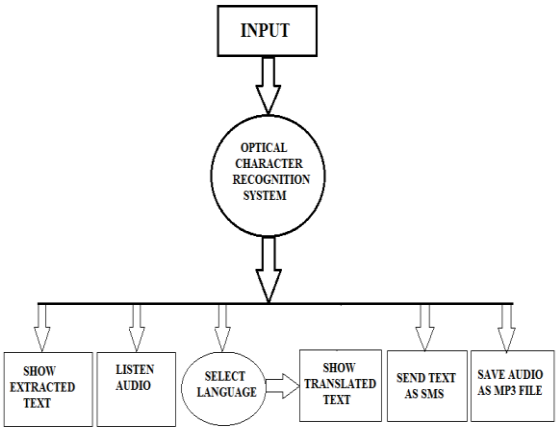

Fig. 1. System concept

2. Related work

There are number of approaches have been proposed for text recognition in natural scenes which can be classified broadly to three categories. Firstly, approaches that can recognize the segmented text by proposing their own features with classifiers training which works well for specific languages and specific data [1],secondly approaches that can recognize and binarize the text without segmentation of text lines using multiple hypothesis frames work [2], [3], thirdly approaches that can improve recognition rate by enhancing the text through binarization which works well for carved text which may have complex background and low resolution [4]-[6].For learning features from unlabeled data many approaches have been proposed [7-9]. Authors in [10] have shown that performance

can grow with large numbers of low level features while in [11], authors have shown that performance can grow with large number of high level features. Text extraction techniques are generally studied because text inserted in images and videos provides important information. There are many feature of text regions have been summarized and characterized effectively by several features. Currently, several OCR systems are available commercially on the market. Some of those are ABBYY FineReader, Microsoft Office Document Imaging and OmniPage. Several more systems such as GOCR, OCRAD and Tesseract OCROPUS are offered by the open source and research communities. However, in most of the systems mentioned above, all the operations is implemented at the backend server and not in real time. To work properly Thus Internet connection is must for these apps. As the application has to communicate with backend server, to get results response time should be increased. In this proposed system all operation are performed in the mobile phone itself, and there is no need of back end server. The application can work as intentional without connection of internet. However, for the text translation to work properly, connection of internet is required.

3. System flow

In our propose system following are the steps in text recognition and translation algorithm.

1. Morphological edge detection
2. Feature filtering
3. Binarization
4. OCR
5. Correction
6. Translation
7. Display of the translation
8. Text-to-speech

The system can use any type of mobile phone camera that points directly towards, the natural scene in order to detect and recognize, translate and overlay text embedded on natural images.

A. Assumptions and constraints

For best results, the images need to meet certain requirements:

Resolution: High-resolution camera image works best. It is recommended that each line of text in the image to be of at least 20 pixels height.

Orientation: Only image with horizontal left-to-right with in 20 degree skew/orientation text are recognized. Proposed method may not work well with images containing text skew in excess of 20 degrees clockwise and anticlockwise.

Languages, fonts and character sets: Proposed method is designed to work on printed, on connected, non-italic, non-underlined text.

Image quality: Sharpness and the contrast of the images should be good. Blur as well as bad camera focus will decrease the quality of the detected text.

Camera Position: Camera position with respect to image must be nearly parallel. Page Layout: Proposed method

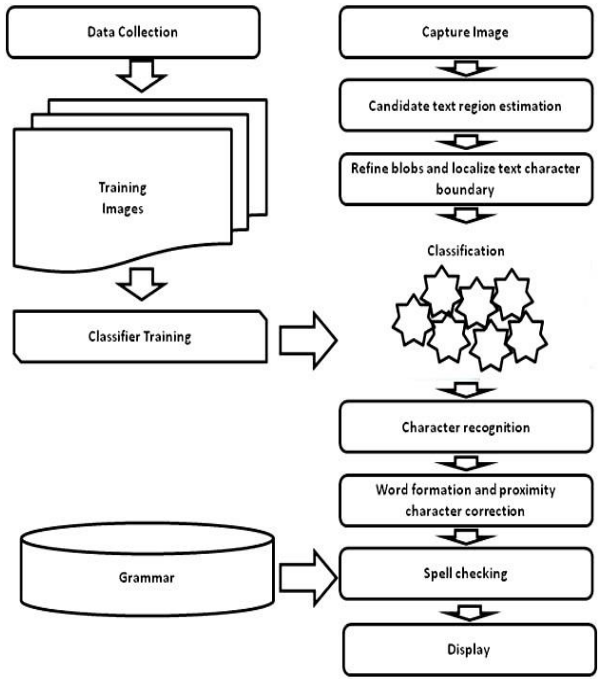

Fig. 2. Flowchart of the work

4. Proposed methodology

Methods for text extraction and segmentation. These methods are as follows:

1) RGB to Grayscale Transformation
2) Thresholding
3) Sobel Edge Sensing
4) Segmentation
5) Template Matching

A. Step 1-Morphological edge recognition

To implement the edge detection algorithm, we first transform the input RGB color image to an image of gray-scale intensity. Y using (1), where red, green and blue components of the input image represented by R, G, and B.

$$\Psi=0{:}229P+0.587\Gamma+0{:}114B \quad (1)$$

To reduce false edge noise and over-segmentation the gray-scale image is then blurred using open-close and close-open filters. Structuring element used for such operation is a 3 X 3 8-connected element. Next, a morphological gradient recognition operation is performed on the blurred image Ybl, as shown in (2).

$$\Psi 2 = M\Gamma(\Psi\beta\lambda;B) = \delta\iota\lambda\alpha\tau\iota o\nu(\Psi\beta\lambda)\aleph\varepsilon\rho o\sigma\iota o\nu(\Psi\beta\lambda) \quad (2)$$

In order to get the threshold level of Y2, we use a global non histogram-based thresholding method. The threshold level is

determined by (3), where s is an edge detector obtained by applying central difference edge detection filter to Y2. [1]

$$\gamma = \frac{\sum Y_2 \cdot s}{\sum s} \tag{3}$$

B. Step 2 - Text feature filtering

To reduce the number of connected components that have to be analyzed, to the binary edge image obtained from Step 1 on which a close operation with a 5 by 5 structuring element is implemented.

After that close operation, all connected components of the edge image are screened with their area, position and size information. A candidate of letter should meet a set of constraints in shape and size. In our approach, connected components we select as letter candidates

After the first round filtering, maximum number of the non-letter components would be removed. So the majority of the left candidates should be letters with the same size and font. Based on this status, we compute the mean height hm of the bounding box of the remaining objects(words) and remove any connected objects with its height smaller than 0:6hm or greater than 1:8hm.

C. Step 3 - Text area binarization

To the original gray-scale image each remaining bounding box is used as a mask. Otsu's method [2] is used to get the threshold values of the masked gray-scale image for binarization. Since each bounding box is relatively small compared to the entire image size, there is no further adaptive thresholding technique is performed. After this step, as the foreground only stroked letters are left, 1, and the rest of the image would go to background, 0.

D. Step 4-6 - Text detection, correction, and translation

The project is focused on implementing text extraction on a mobile phone, we implemented the next three steps - text detection, correction and translation on a server with open source software for the sake of simplicity's. As the optical text recognition engine we using Google's open source OCR – Tesseract [3]. To perform text correction Peter Norvig's algorithm [4] is added to the routine. Then to translate the text into different language the Google translator [5] is used.

E. Step 7- Display of the translation

From the server the translated text string from Step 6 is sent to the mobile device (Andriod phone), and then displayed at the screens top center region.

5. Algorithm

The steps used for converting a scanned document to text are:

1) Load template images.
2) Load the scanned image of the document which is converted to text.
3) Transform the scanned image to gray scale.
4) To remove dust use a low-pass Finite Impulse Response (FIR) filter for filter the scanned image.
5) Whitespace in the text lines, break the document into number of lines of text.
6) Based on whitespace between the characters Break each line into characters; using the average character width, determine where spaces occur within the line.
7) For each character, decide the most closely matching character from the training images and attach that to the resultant text; for each space, attach a space character to the resultant text.
8) Output the increased text.
9) Return to step 2. If there are any more scanned images to be transformed to text.

6. Data flow diagrams

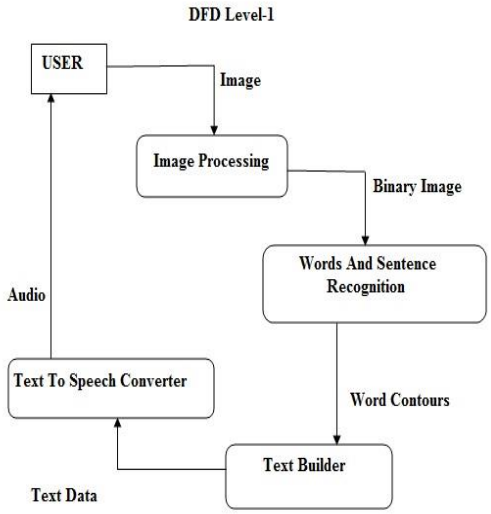

Fig. 4. DFD Level-1

7. Conclusion

Our paper presented an application for performing Text recognition and Translation on images captured by Android Smartphone. This android application has the ability to generate excellent results. Research show that the application performs very well for recognition of phrase and individual character. The application Overall accuracy can be improved by using manual thresholding, segmentation and classification throughout the recognition process. Moreover, we have presented application how works and how different techniques are performed. We transform RGB image into Grayscale image, then we transform grayscale image into Binary image, then we implement edge recogntion, template matching and segmentation to produce extracted text from image. This proposed system is efficient as well as very less memory consumes. It worked efficiently with devices having 512 MB RAM.

A. Advantages

1) Able to work offline. No need of Connectivity of Internet.
2) In very short amount of time.
3) Convenient as mobile devices are portable.

B. Limitations

1) Accuracy rate is straightly dependent on input image quality.
2) Cursive characters difficult to recognize.
3) Strictly Android platform.

Appendix

OCR – Optical Character Recognition
TTS – Text To Speech
FIR – Finite Impulse Response
PDA – Personal Digital Assistant

Acknowledgement

The authors would like to thank Department of Computer Engineering, SIT Lonavala and indebted to our guide Prof. S. B. Nimbekar for his guidance and sagacity without which this paper would not have been designed. He provided us with valuable advice which helped us to accomplish the design of this paper. We are also thankful to our HOD Prof. S. D. Babar (Department of Computer Engineering) for his constant encouragement and moral support. Also we would like to appreciate the support and encouragement of our colleagues who helped us in correcting our mistakes and proceeding further to produce the paper with the required standards.

References

[1] Chucai Yi,YingLi Tian and Arditi, A.Portable Camera-Based Assistive Text and Product Label Reading From Hand-Held Objects for Blind Persons IEEE, May 2013.
[2] Derek Ma, Qiuhau Lin, Tong Zhang ―"Mobile Camera Based Text Detection and Translation" Stanford University, Nov 2000.
[3] Yassin M. Y. Hasan and Lina J. Karam, Morphological Text Extraction from Images. IEEE Transaction on Image Processing Vol. 9 no.11, Nov 2000.
[4] K. Takeda, K. Kise, and M. Iwamura, "Memory reduction for real-time document image retrieval with a 20 million pages database," Proceedings of the 4th International Workshop on Camera-Based Document Analysis and Recognition, 2011.
[5] Snehal Charjan, R. V. Mante, P. N. Chatur,"Character localization and recognition application for Smartphone," International Journal of Advanced Computer Research, vol. 2, Dec. 2012.
[6] Lukas Neumann, Jiri Matas ,Text Localization in Real-world Images using Efficiently Pruned Exhaustive Search, International Conference on Document Analysis and Recognition, 2011.
[7] Rong Huang, Shinpei Oba, Shivakumara Palaiahnakote and Seiichi Uchida, Scene Character Detection and Recognition Based on Multiple Hypotheses Framework, 21st International Conference on Pattern Recognition (ICPR 2012) November 11-15, 2012.
[8] Saumya Jetley, Swapnil Belhe, V.K. Koppula, Atul Negi,Two-Stage Hybrid Binarization around Fringe Map based Text Line Segmentation for Document Images, 21st International Conference on Pattern Recognition (ICPR 2012), November 11-15, 2012.
[9] Chucai Yi, Yingli Tian and Aries Arditi ,Portable Camera-Based Assistive Text and Product Label Reading From Hand-Held Objects for Blind Persons, IEEE/ASME Transactions on Mechatronics, vol. 19, no. 3, June 2014.