Design of Text Detection and Translation System for Camera Based Android Smartphone

Rahul B. Panchal[1] Rutuja G. Sonawane[2] Himayat Shaikh[3] Piyush P. Gawali[4]
[1,2,3]Student [4]Assistant Professor
[1,2,3,4]Department of Information Technology
[1,2,3,4]NBN Sinhgad School of Engineering, Pune

Abstract— In modern technology, the field of digital image processing has grown up to a new dimension due to availability of high resolution cameras. As the technology is growing rapidly, various technologies are being developed for handheld devices like Smartphone and tablets. In recent years, Android platform has gained huge popularity in terms of number of available applications and market share. The goal of our project is to design Android-platform based text detection and translation application that is able to recognize text present in digital image, translate the text from English to various other languages like Arabic, German, French, Russian, Spanish etc. As well as convert the text into speech and allows the audio file to be stored locally for later use. Experimenting with a set of 100 random business card pictures, captured by Smartphone camera, we have achieved a maximum accuracy-rate that is greater than 90%. Moreover, the developed system is more efficient and consumes less memory so as to be suitable for Smartphone devices.

Key words: OCR, Text Detection, Text Translation, Android, text-to-speech

I. INTRODUCTION

Optical character recognition is abbreviated as OCR. It is powerful tool for digitizing the information present in our analog lives. This technology has been used in past for building digital libraries, understanding hand-written office forms, recognizing text from natural scenes, and etc. With the use of OCR technology, camera-captured or scanned documents can be converted into machine editable soft copies that can be reproduced, edited, searched and transported with ease. Our project's aim is to enable OCR on Smartphone. In this application OCR technology enables the conversion of captured or scanned images of printed text into information or text that can be understood or edited using android Smartphone.

In modern technology, Growing computational power of portable mobile devices and the availability of cheap digital cameras on them makes it possible to switch from traditional hand dictionary translation to a new comfortable, faster and affordable way. Few years ago, research in the field of OCR that is Optical Character Recognition was limited to document images captured with flatbed desktop scanners. These devices are not portable because of large size of scanners and thus usability of such system is limited. Recently, with the advancement of internal memory and processing speed of hand-held mobile devices like iPhone, Smartphone, iPods, Personal Digital Assistant (PDA) etc. Having built in digital cameras, a new direction of research has emerged into picture. Today, mobile phones are one of the most commonly used electronic devices. Mobile phones with high resolution cameras (above 3 megapixel), powerful processors (above 1GHz) and a range of different embedded sensors (GPS, accelerometer, compass) are widely deployed and becoming popular. By fully utilizing these advancements, mobile phones are becoming powerful portable computing platform. They can process computing-intensive applications in real time. We have developed an android application that can detect text information within an image captured with mobile phone camera, extracts it, recognizes it and translate it from English to some other languages like Arabic, German, French, Russian, Spanish etc. This android application can be very helpful for tourists travelling in foreign environment.

We believe that this mobile solution to bring information from physical world into digital world can be a good match for future, but camera-captured documents have some drawbacks due to non-contact nature of digital cameras attached to mobile devices. Camera-captured images often suffer a lot from insufficient and uneven document lighting, focus loss, and geometrical distortions such as text skew, text misalignment and bad orientation. Since our application is running on mobile phone, real time response is also a serious issue. There is open source free software for OCR known as Tesseract, which is considered as most accurate OCR software in existence.

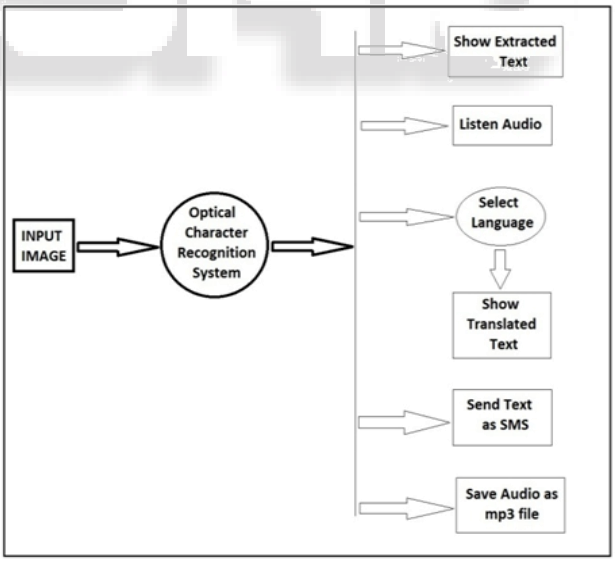

Fig. 1: System Concept

II. RELATED WORK

Currently, there are various OCR systems available commercially on the market. Some of those are OmniPage, ABBYY FineReader and Microsoft Office Document Imaging. Additionally, several more systems like OCROPUS, OCRAD, GOCR and Tesseract are offered by the research and open source communities. However, in most of the systems mentioned above, all the processing is performed at the backend server and not in real time. Thus

Internet connection is must for these apps to work properly. As the application needs to communicate with backend server, the response time to get results also increases. In this project, we have developed an Android application which performs all the processing in the mobile phone itself and there is no need of backend server. The application can work as intended without internet connection. However, for the text translation to work properly, Internet connection is required.

Several commercial products and academic projects have tried to exploit Smartphone camera to develop interesting applications. There is an application for iPhone named 'Word Lens' which translates text visible in viewfinder and overlays the result on the screen in real time. Word Lens application is also capable of working without internet connection. For Smartphone running on 'Windows Phone 7', there is an application named 'Snap and Translate' which also uses OCR technology. In this application, consider a person is reading English article and he wants Chinese translation of some English word/phrase or paragraph. So here that person can use a windows phone to snap a picture of the text, tap the word or swipe the phrase or circle the paragraph with a finger, and he will get a Chinese translation displayed on the screen of the phone.

III. PROPOSED METHODOLOGIES

In this paper, we propose several methods for text extraction and segmentation. These methods are as follows:

1) RGB to Grayscale Conversion
2) Thresholding
3) Sobel Edge Detection
4) Segmentation
5) Template Matching

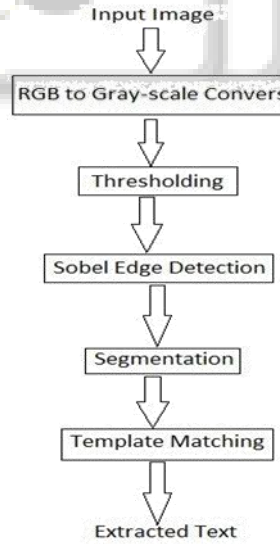

Fig. 2: System Flow

A. *RGB to Grayscale Conversion:*

The very first step is we convert input RGB image to Grayscale image using expression given below, where R, G, B represents red, green and blue components respectively of input image.

$$Y = 0.299R + 0.587G + 0.114B$$

Grayscale image is nothing but an image in which each pixel value is a single sample which contains intensity information only. Such images are also known as Black and White images. The intensity of each pixel is expressed between minimum and maximum range. Grayscale images are also referred as monochromatic which means the presence of only one color that is mono chrome.

B. *Thresholding:*

Thresholding is referred as simplest method of image segmentation. To create binary image from Grayscale image, Thresholding is used. In the process of thresholding, each individual pixel in an image is marked as object pixel if its value is greater than some threshold value and if its value is lower than threshold value then it is marked as background pixel. This convention is referred as threshold above. There is another variant named threshold below, which is exactly opposite of threshold above; threshold inside, where a pixel is labelled as object if its value is between two thresholds; and threshold outside, which is exactly opposite of threshold inside. Generally, an object pixel is given a value of '1' and background pixel is given a value of '0'. In the end, by coloring each pixel white or black, depending on pixel's label, a binary image is created.

C. *Sobel Edge Detection:*

1) *Blurring:*

When we have binary image as input, next step is blurring. In process of blurring, each pixel in source image gets spread over and mixed into surrounding pixels. In the destination image, each blurring pixel is made up out of a mixture of surrounding pixels from source image. Blurring process reduces the sharpening effect of image, but it makes text detection more accurate.

2) *Thinning:*

There are various thinning algorithms to implement process of thinning. Thinning process performs morphological operation which is used to remove foreground pixel from binary image. We are using Stentiford Thinning Algorithm to implement thinning process. In this algorithm, four set of 3*3 template is used for scanning image.

This is how these templates are used:

In this, our pixel of image is matched with central pixel of four templates. Ignore the pixel, if it doesn't match with any template of pixels. Mark the pixel as erasable, if that pixel is not an endpoint pixel. Endpoint pixel has less than two pixel of same color out of eight.

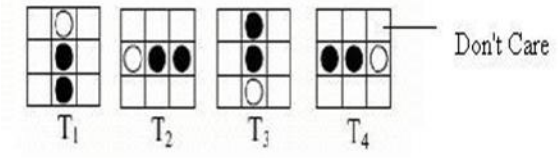

Fig. 3: Pixels are matched with the above template

D. *Segmentation:*

The process of separating character from an image is nothing but segmentation. We can perform further processing on character, when we have each character separately. To separate each character from an image, we use horizontal and vertical scan lines.

1) *Scaling:*

Scaling means process of changing size of an image. After segmentation, when we get box of character, it is in variable dimension. Thus we need to bring box of character in standard size which is 100*100 dimensions for further

processing. This dimension depends on our template dimension which is present in our database. There are various algorithms to implement scaling. We are using nearest neighbour interpolation algorithm for scaling here. In this algorithm every pixel is replaced with four pixels of same color.

E. Template Matching:

Template matching is one of the important techniques used in image processing. In template matching, we match our generated template of character with the available template of character which is stored in our database. While comparing both templates, we maintain score of each template. In the end, we search for maximum score of template and display its respective character as resulted character. In similar way, we get each set of character.

IV. TEXT TRANSLATION AND TEXT TO SPEECH

We are using Google Translator for translating the text into several other languages such as Arabic, German, French, Chinese, Dutch, Russian, Spanish and many more. The translated text is sent back to mobile phone i.e. Android Smartphone from server and then displayed on screen.

In android, there is an text-to-speech(TTS) synthesizer. As name suggests it can read text aloud automatically. The text which is extracted from Optical Character Recognition is given to TTS synthesizer. TTS synthesizer can be implemented by both software and hardware. The artificial production of human speech means speech synthesis. A text-to-speech synthesizer system can convert normal language text to speech.

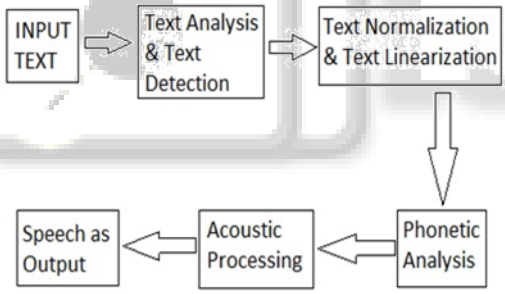

Fig. 4: Text to Speech Conversion

V. GENERAL ALGORITHM

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

1) Load template images.
2) Load the scanned image of the document to be converted to text.
3) Convert the scanned image to gray scale.
4) Filter the scanned image using a low-pass Finite Impulse Response (FIR) filter to remove dust.
5) Break the document into lines of text, based on whitespace between the text lines.
6) Break each line into characters, based on whitespace between the characters; using the average character width, determine where spaces occur within the line.
7) For each character, determine the most closely matching character from the training images and append that to the output text; for each space, append a space character to the output text.
8) Output the accumulated text.
9) If there are any more scanned images to be converted to text, return to step 2.

VI. DATA FLOW DIAGRAMS

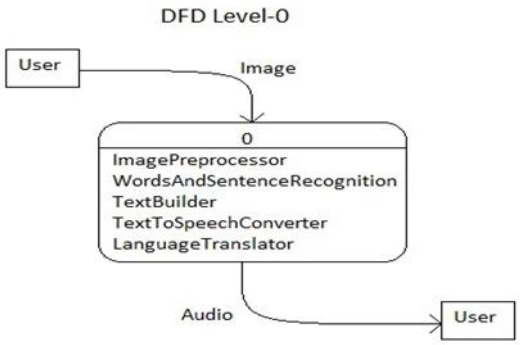

Fig. 5: DFD Level 0

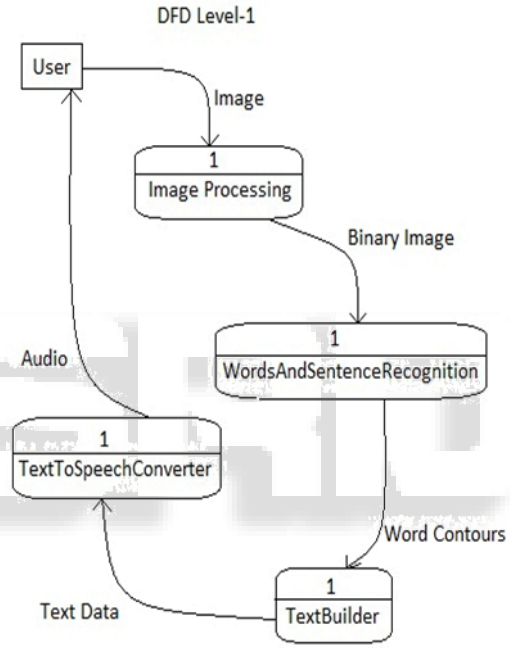

Fig. 6: DFD Level 1

VII. CONCLUSION

This paper presented a application for performing Text Detection and Translation on images captured by Android Smartphone. The application has the ability to produce excellent results. Experiments show that the application performs very well for individual character and phrase recognition. Overall accuracy of application can be improved by using manual segmentation, thresholding and classification throughout the recognition process. Furthermore, we have presented how application works and how different techniques are implemented. Firstly, we convert RGB image into Grayscale image, then we convert grayscale image into Binary image, then we perform edge detection, segmentation and template matching to produce extracted text from image. The application is efficient and consumed very less memory. It worked well with devices having 512 MB Ram.

A. Advantages:

1) Able to work offline. No need of Internet Connectivity. (Only text translation feature requires internet.)
2) Convenient as mobile devices are portable.
3) Very low memory consumption.
4) Produce results in very short amount of time as there is no need of backend server.

B. Limitations:

1) Accuracy rate is directly dependant on quality of input image.
2) Strictly Android platform.
3) Difficult to recognize cursive characters.

VIII. APPENDIX

1) OCR – Optical Character Recognition
2) TTS – Text To Speech
3) FIR – Finite Impulse Response
4) PDA – Personal Digital Assistant

IX. ACKNOWLEDGMENT

The authors would like to thank Department of Information Technology, NBNSSOE and indebted to our guide Prof. Piyush P. Gawali 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. P. Patil (Department of Information Technology) 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] 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.
[2] Ariffin Abdul Muthalib, AnasAbdelsatar, Mohammad Salameh, and JuhriyansyahDalle, ―"Making Learning Ubiquitous With Mobile Translator Using Optical Character Recognition (OCR)", ISBN:978-979-1421-11-9.ICACSIS 2011.
[3] Snehal Charjan, Prof. R. V. Mante, Dr. P. N. Chatur,―"Character localization and recognition application for Smartphone",International Journal of Advanced Computer Research, vol-2 Dec-12.
[4] X. Liu and D. Doermann, "Mobile retriever: access to digital documents from their physical source," *Int. J. Doc. Anal. Recognit.*, vol. 11, pp. 19–27, September 2008.
[5] Derek Ma, Qiuhau Lin, Tong Zhang ―"Mobile Camera Based Text Detection and Translation" Stanford University, Nov 2000.
[6] Nitin Mishra,C Patvardhan, ―"ATMA: Android Travel Mate Application", International Journal of Computer Applications (0975 – 8887) vol 50 – No.16, July 2012.
[7] The Tesseract open source OCR engine, http://code.google.com/p/tesseract-ocr.
[8] Abbyy Mobile OCR Engine. http://www.abbyy.com/mobileocr/.
[9] GOCR - A Free Optical Character Recognition Program. http://jocr.sourceforge.net/.
[10] OCRAD - The GNU OCR. http://www.gnu.org/software/ocrad/.
[11] OCRopus - Open Source Document Analysis and OCR System. https://github.com/tmbdev/ocropy
[12] Wikipedia OCR details. http://en.wikipedia.org/wiki/Opticalcharacter recognition.
[13] Simple OCR - Optical Character Recognition. http://www.simpleocr.com/.
[14] D. Chen, J.-M. Odobez, and H. Bourlard. Text detection and recognition in images and video frames. Pattern Recognition, 37(3):595 – 608, 2004.
[15] B. Epshtein, E. Ofek, and Y.Wexler. Detecting text in natural scenes with stroke width transform. In IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pages 2963–2970, June 2010.

