IJCSE International Journal of Computer Sciences and Engineering Open Access
Survey Paper Vol.-7, Issue-4, April 2019 E-ISSN: 2347-2693

Mobile Based OCR Systems: State-of-the-art Survey for Indian Scripts

Ravneet Kaur[1*], Dharam Veer Sharma[2]

[1] Dept. of Computer Science, DAV College, Jalandhar, Punjab, India
[2] Dept. of Computer Science, Punjabi University, Patiala, Punjab, India

[*]*Corresponding Author: ravneet@davjalandhar.com*

DOI: https://doi.org/10.26438/ijcse/v7i4.457461 | Available online at: www.ijcseonline.org

Accepted: 13/Apr/2019, Published: 30/Apr/2019

Abstract— Few decades ago, approach to character recognition was limited to desktop scanner. The usability of such system was limited as they were non portable because of large size. With the advent of technology and portable computing devices such as mobile phone, PDA, iPhone etc. new trends of research has emerged, where Mobile phones are the most commonly used electronic device, eliminating the need for bulky devices like scanners, desktops and laptops. The convergence of powerful processors and high resolution cameras on mobile devices has directed the focus of research to development of mobile applications, where image processing applications such as OCR's are in demand. This paper present State-of-the-Art survey of Character Recognition systems for mobile devices and summarize some commercially available OCR applications.

Keywords— *Mobile OCR, Document Image Processing, Text Recognition.*

I. INTRODUCTION

An image is an artifact that depicts visual perception. Humans are able to see images because of the communication between eyes and brain where eyes act as an input for brain. The ability to perceive these visualized images varies from person to person. Similarly, we have OCR technology in the field of computer through which textual information contained in images is converted into editable text format. It enable users to convert different documents such as scanned paper documents, PDF file or images into machine editable soft copies.

By 2018, global mobile usage is more than 4.5 billion, which is around 65 percent of the world population. Mobile internet usage takes over desktop internet usage as more than half the world now uses a Smartphone ; the reason for ever increasing demand for mobile phone is its 'anywhere, anytime' feature. This technology has long seen use in building digital libraries, recognizing text from natural scenes, understanding hand-written office forms etc. By applying OCR technologies, scanned or camera captured documents are converted into soft copies that can be edited, searched, reproduced and transported with ease.

Mobile based OCR systems reduce user input, especially keyboard input; the ideal case is "key-press-free" input. In addition such systems facilitate mobile users to process short segments of text on demand with the expectation of low response time. In brief, the paper is organized as follows, Section I contains introduction to mobile based OCR systems , Section II highlights related work in the area, In section III various challenges related to Image Processing using portable digital imaging devices are given, Section IV describes results and discussion. Section V concludes research work with future directions.

II. RELATED WORK

Document image processing and understanding has been extensively studied over the past fifty years. Work in the field covers many different areas including document segmentation, text extraction, text recognition, enhancement, graphics analysis, handwriting recognition, form processing, signature verification etc. In addition, the increasing availability of high performance, low priced, portable digital imaging devices has created an opportunity for on demand analysis of documents as they can capture images of any kind of document including very thick books, historical pages too fragile to touch, and text in scenes.

Doermann et al. [1] presented a survey of application domains, technical challenges and solutions for recognizing documents captured by digital cameras. **Ma et al. [2]** developed an android-platform based text translation application that recognize English text captured by mobile phone camera and translate the text. Text extraction and recognition rate greater than 85% on character level is reported. **Luo et al. [3]** presented camera based mobile OCR systems for business card images. Text region in image are first skew corrected and binarized thereafter. Such text

regions are segmented into lines and characters, and subsequently passed to an OCR engine for recognition. The OCR engine is designed as a two layer template based classifier. A similar system is presented for Chinese-English mixed script business card images in **Luo et al. [4]**. A system for English capital letters is developed by **Laine et al. [5]**. The captured image is first skew corrected by looking for a line having the highest number of consecutive white pixels and by maximizing the given alignment criterion. Then, the image is segmented and recognized. The accuracy reported is not satisfactory for real life applications. **Liang et al. [6]** presented a framework for restoring the frontal-flat view of a document from a single camera-captured image. Framework provides a unified solution for both planar and curved documents and can be applied in many, especially mobile, camera-based document analysis applications.

Tesseract OCR engine developed by Hewlett-Packard is a free software for various operating systems. Currently it is maintained by Google and can convert images to text in over 60 languages. Many researchers [7-12] worked for improving the accuracy of tesseract OCR engine. **Smith et al. [7]** in his work trained the classifier for English language on a mere 20 samples of 94 characters from 8 fonts in a single size, with 4 attributes: normal, bold, italic, bold italic making a total of 60160 training samples. **Hasnat et al. [8]** have trained Tesseract OCR engine for Bangla script, and observed the results using scripts of different sizes and fonts. Recognition accuracy of 97% at character level and 88% at word level is reported. **Mahbub et al. [9]** developed mobile camera based Bangla text detection and translation application on Android-platform. Application recognizes the Bangla text at word level, converts text into editable Unicode text and further translates Unicode text into English. They have used OCR engine Tesseract, Google translate API and an open source Android application called android-ocr for development. Application works comparatively better for large & bold text than the small and normal text. **Mishra et al. [10]** presented methodology to improve recognition accuracy of Tesseract OCR engine for Hindi Language. In the Pre-processing Phase, the horizontal and vertical histograms are generated for each line of the text identified in the test image. Sufficient improvement in the recognition rate, processing time and the size of training database is reported after integrating Shirorekha Chopping with Tesseract OCR Engine. Comparison with other Devanagari OCR engines on the basis of recognition accuracy, processing time, font variations and database size is also given. **Badla et al. [11]** worked on Tesseract OCR for improving the efficiency for Hindi language for Mobile devices. Tesseract is trained for Hindi character set in 4 fonts. Pre processing steps Rotation, optimize resolution, adjust the DPI's and lastly the Luminosity grayscaling algorithm are applied to improve the recognition of the characters. Average runtime of 681 ms is improved to 459 ms after pre processing by keeping the accuracy almost the same. **Chowdhury et al. [12]** build an Android application that can extract text from any image that contains Bengali characters and convert it into an editable document. Tesseract was used to recognize the characters which utilizes Leptonica Image Processing library to process image and extracting data from the image. To recognize more characters and joint letters, they worked on decreasing the rate of error to preserve more texts. Joint letters, dangerous ambiguity and contrast issues were handled to increase efficiency. An average precision of 66.8% and recall of 70.9% for the desktop and precision 65.5% and recall 70.0% for mobile version is reported.

Chang et al. [13] focussed on implementing pre-processing Techniques for OCR Applications on Mobile Devices using desktop open sources library OpenCV. **Mollah et al. [14]** designed a complete OCR system for camera captured business cards for English language. Text regions are extracted, skew corrected, binarized and segmented into lines and characters. Characters are passed into the recognition module. Experimenting with a set of 100 business card images, captured by cell phone camera, a maximum recognition accuracy of 92.74% is reported. Compared to Tesseract, presented recognition accuracy is worth contributing. Moreover, the developed technique is computationally efficient and consumes low memory so as to be applicable on handheld devices.

Nguyen et al. [15] developed an automatic number-plate recognition application that utilizes the OpenCV library for locating the number plate area and then uses Kohonen neural network to recognize characters. Method achieved average recognition accuracy of 90% with fast processing speed. **Gosavi et al. [16]** developed Mobile OCR Application for English language Using Neural Network. Kohonen algorithm is used for reorganization of characters which produce binarized image as output. Also translation of recognized English text into Marathi language is performed. **Chandrashekhar et al. [17]** worked for Kannda characters recognition for handheld devices using neural network. Input image is converted into gray scale which then recognized by system. English translation of recognized Kannada word is displayed along with the voice output.

Cutter et al. [18] developed a system, implemented as an iOS app, enables two interaction modalities autoshoot and guidance. System autoshots and acquire a well-framed image to help blind users of a mobile OCR app to take OCR-readable pictures faster.

III. CHALLENGES IN THE DEVELOPMENT

The Development of OCR systems for handheld devices is more challenging as compared to desktop systems. The non-contact nature of digital camera and unavailability of sophisticated focussing system make camera captured images

to suffer from environmental conditions. One challenge in adapting OCR for mobile devices is in handling lower quality input images. A second challenge in running OCR on mobile devices is same as running OCR for desktop, in particular recognition of complex scripts and degradations of documents.

Device and Scene Related Challenges:

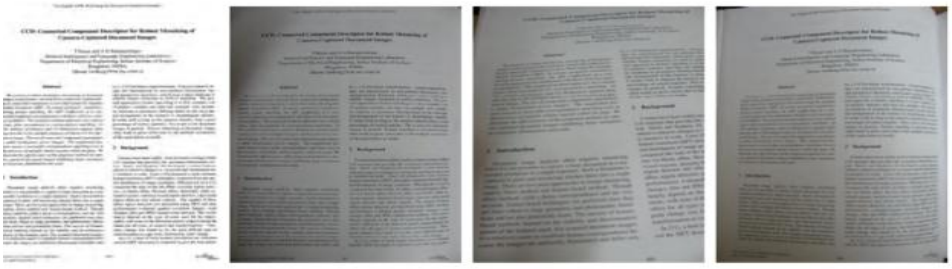
Recognition of text under handheld devices is different from flatbed scanners. Camera captured images very often suffer from environmental conditions, such as insufficient and uneven document lighting, reflection, out of focus and geometrical distortions such as text skew, text misalignment, bad orientation and scaling. Further, since the system is for mobile phone, real time response is also a critical challenge.

i. **The Moving Camera**: 1:1(one-to-one) mapping of image and the document is not always possible. Unlike a camera mounted on a tripod or surveillance camera mounted on a wall, a camera phone is typically in motion. Due to variation in distance and inclination of the camera from the document, the capture conditions may result in motion blur, out of focus blur and low resolution.

ii. **Uneven illumination and Shadows**: Unlike scanners, digital cameras have far less control of the lighting conditions of the acquisition environment. Uneven and insufficient illumination makes certain parts of the image abnormally brighter than the rest of the image, In particular the centre of the view is bright and the intensity gradually decays towards the image boundary. The background image may also contain some shadows due to environmental factors. Uneven lighting and irrelevant shadows present challenges to text recognition system. Care should be taken while capturing the image with very less shadow falling on the image. The processing and result of shadows may vary from device to device depending on the device configuration.

iii. **Perspective distortion**: Images acquired by handheld device very often suffer from skew and perspective distortion. Perspective distortion occurs when the text plane is not parallel to the imaging plane. The effect is that characters situated farther away appear smaller and the parallel line assumption no longer holds in the image. This results in increased difficulty in performing layout analysis and decomposition of the document image into text lines, words and characters. In addition, the characters get deformed leading to difficulty in recognition. The conventional process of skew detection and correction cannot handle perspective distortion. Figure 1: Issues with Camera-Captured Images. (**a) A Scanned Document Image. Camera captured images exhibiting (b) Uneven Lightning (c) Perspective Distortion (d) 3D Deformation.**

iv. **Warped text**: Pages of an opened book are usually curled and are rarely flat. Image warping and rendering techniques are required to handle geometric distortions. In any camera-based document analysis system, it is first required to normalize the photometric and geometric distortions present in the acquired image and render it into a form similar to that of a scanned image. Current document analysis systems fail even under a moderate warping.



(a) (b) (c) (d)

Figure 1: Issues with Camera-Captured Images. (a) A Scanned Document Image. Camera captured images exhibiting (b) Uneven Lightning (c) Perspective Distortion (d) 3D Deformation.

v. **Lack of Floating Point Units (FPU):** In addition to the above challenges, smart phones also do not have a Floating Point Unit which is required for floating point arithmetic operations. Floating point emulators are used to perform arithmetic operations on such devices but that result in slower operation.

vi. **Computational Power and Battery Life:** Though the processing speed and memory size of handheld devices have improved in recent times but it is still not sufficient enough to run desktop based OCR algorithms as they are computationally expensive and require very high amount of memory. Also multiple tasks may run concurrently. From the software development point of view, we cannot do much about the system and hardware architecture. One possible solution is to improve the algorithm's efficiency, by avoiding floating point operation, using approximated algorithms.

vii. **Complex Background:** Scene text can appear on any surface, not necessarily on plane surface. Scene text understanding normally involves text detection and extraction before character recognition task. The recognition task is performed only on the detected text regions so as to mitigate the effect of background complexity. **Error! Reference source not found.** highlights these issues that arise in natural scene images.

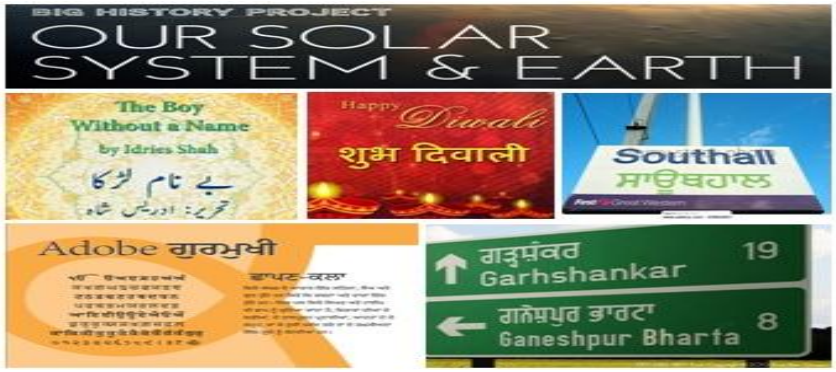


Figure 2: Scene images exhibiting presence of arbitrary text orientation, perspective distortion, multi-script content, multiple colors, complex backgrounds and artistic font styles

Document-Related Challenges:

The quality of a document image depends on the distortions present in the document. Degradation in document images such as poor quality of paper, the printing process, ink blot or fading, document aging, extraneous marks, wrapping, coping noise, merged/broken characters, variation in font style and size, language complexity etc drastically affect the performance of OCR system. Document image distortions can be categorized as follows:

i. **Poor Quality Document:** Low quality of old documents: poor paper quality, aged, faded causes a lot of errors in the digitization process. So pre-processing of the image is required before proceeding to other stages. This involves binarization of images, orientation, skew and slant correction and noise removal of images.

ii. **Document Segmentation**: Layout analysis is performed to automatically extract image component. The major components of document image are text, pictures, headings, tables, horizontal and vertical lines. Purpose of document segmentation is to identify all individual components, build suitable digital representations and automatically recognize the contents of Text Region using suitable optical character recognition techniques. Presence of various components, variations in fonts shape and sizes, and multiple text regions in same document image make the task challenging.

iii. **Touching Character Problem**: This is the most common problem when we talk about Indian scripts such as: Devnagari, Gurmukhi, Bnagla. In this category of degraded text, two neighbouring characters touch each other. Different categories are: a) Characters in upper zone of a line touching with each other. b) Characters of upper and lower zone of text touched with the middle zone. c) Character in lower zone of a line touching with each other.

iv. **Variations in Text:** Classifier may show poor results due to the presence of a large variation in font shapes and sizes in the document. Handling a document that contains different size characters is also challenging.

v. **Line, word and character Segmentation**: Line segmentation is challenging when touching lines are present in the document. Word segmentation is challenging task due to variations in space between words in particular the segmentation of tightly spaced words is very difficult to handle. Character Segmentation is also very challenging as characters in a word may get touched with each other. The biggest challenge involved in recognition of touching characters is to identify the position at which the touching pair of characters must be segmented.

vi. **Multiple Scripts**: In a multi-lingual country like India, many documents, forms and signboards are generally bi-lingual or tri-lingual in nature. It is common to find English words interspersed within sentences in Indic-script documents. Every script has certain characteristics that distinguish it from other scripts and may require script-specific processing methods. Identification of script plays a vital role for the development of next generation OCRs.

IV. RESULTS AND DISCUSSIONS

Both device and document related challenges make the task of Character Recognition difficult. Noise in the input image contribute to decrease in accuracy of OCR system. It is either

due to the quality of paper on which printing is done or due to capturing conditions. State-of-the-art OCR systems can produce good results for high resolution, high quality document images with a clean background for foreign and national languages, but very few systems are available to recognize Indian languages. Therefore, the need of the hour is to design computationally efficient and light-weight OCR algorithms that can run smoothly on handheld mobile devices.

V. CONCLUSION

With widespread applications of computer and multimedia technologies, demand in creating paper less environment is increasing. Automatic procurement from documents become an active research area, much knowledge is acquired from documents such as reports, government files, books, journals, newspapers, magazines, letter etc. The manual acquisition of text from such documents can involve an extensive amount of handcrafting. Such handcrafting is time consuming and very laborious. A very few mobile applications are available to recognise Indian scripts. There is great need to develop Mobile base OCR systems for Indian languages to offer advantage of portability and ubiquity.

REFERENCES

[1] D. Doermann, J. Liang and H. Li,"*Progress in Camera-Based Document Image Analysis*", IEEE International Conference on Document Analysis and Recognition (ICDAR'03), pp. 606-616, 2003.

[2] D. Ma, Q. Lin and T. Zhang," Mobile Camera Based Text Detection and Translation," Stanford University, November 2000.

[3] X. P. Luo, J. Li and L. X. Zhen, "*Design and Implementation of a Card Reader based on build-in Camera*", IEEE International Conference on Pattern Recognition, pp. 417-420, 2004.

[4] X. Luo, L. X. Zhen, G. Peng, J. Li and B. H. Xiao, "*Camera based Mixed-Lingual Card Reader for Mobile Device*", 8th International Conference on Document Analysis and Recognition, pp. 665-669, 2005.

[5] M. Laine and O. S. Nevalainen, "*A Standalone OCR System for Mobile Camera-Phones*", Personal, Indoor and Mobile Radio Communications, IEEE 17th International Symposium, pp. 1-5, September 2006.

[6] J. Liang, D. Menthon and D Doermann, *"Geometric Rectification of Camera-Captured Document Images"*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, issue 4, pp. 591-605, April 2008.

[7] R. Smith, "*An overview of the Tesseract OCR Engine*" IEEE 9th International Conference on Document Analysis and Recognition(ICDAR 2007), Curitiba, Brazil, pp. 629-633, September 2007.

[8] M. A. Hasnat, M. R. Chowdhury and M. Khan, "*Integrating Bangla Script Recognition support in Tesseract OCR*", Proceeding of the Conference on Language & Technology, pp. 108-112, 2009.

[9] S. Mahbub, U. Zaman and T. Islam, "*Application of Augmented Reality: Mobile Camera Based Bangla Text Detection and Translation*", B.Sc.(CSE)- Thesis report, BRAC University, 2012.

[10] N. Mishra, C. Patvardhan, C. V. Lakshmi and S Singh, *"Shirorekha Chopping Integrated Tesseract OCR Engine for Enhanced Hindi Language Recognition*", International Journal of Computer Applications, vol. 39, issue 6, pp. 19-23, February 2012.

[11] S. Badla, "*Improving the efficiency of Tesseract OCR Engine*", Master of Science – Thesis report, San José State University, 2014.

[12] A. Chowdhury, A. Foysal and S. Islam, "*Bangla Character Recognition for Android Devices*" International Journal of Computer Applications, vol. 136, issue 11, pp. 13-19, February 2016.

[13] Loh Zhi Chang, Zhou Zhi Ying and Steven, "*Robust Pre-processing Techniques for OCR Applications on Mobile Devices"*, ACM Proceedings of 6th International Conference on Mobile Technology, Application & Systems, article no. 60, 4 pages, 2009.

[14] A. F. Mollah, N. Majumder, S. Basu and M. Nasipuri, "*Design of an Optical Character Recognition System for Camera based Handheld Devices*", International Journal of Computer Science, vol. 8, issue 4, pp. 283-289, July 2011.

[15]Tuan Nguyen, Don Nguyen, Phu Nguyen," *UIT-ANPR: toward an open framework for automatic number plate recognition on smartphones "*, ACM Proceedings of the 8th International Conference on Ubiquitous Information Management and Communication, January 2014.

[16] M. B Gosavi, I. V Pund, H. V Jadhav and S. R Gedam, " *Mobile Application with Optical Character Recognition Using Neural Network*", International Journal of Computer Science and Mobile Computing, vol. 4, issue 1, pp. 483-489, January 2015.

[17] M. V. Chandrashekhar, M. S Kumar, M. B. N. Taj and K. Asha, "*Optical Character Recognition on the Android Operating System for Kannada Characters using Kohonen Neural Network*", International Journal of Advanced Technology in Engineering and Science, vol. 3, special issue 1, pp. 247-251, May 2015.

[18] M. Cutter and R. Manduchi, "*Improving the Accessibility of Mobile OCR Apps Via Interactive Modalities"*, ACM Transactions on Accessible Computing, Vol. 10, No. 4, Article 11, 27 pages, August 2017.

