KannadaPado: Mobile-based recognition and cross-lingual transcription of camera captured text in Kannada

A. G. Ramakrishnan
Department of Electrical Engineering
Indian Institute of Science
Bengaluru, 560012
Email: agr@iisc.ac.in

Deepak Kumar
Department of Electrical Engineering
Indian Institute of Science
Bengaluru, 560012
Email: dipkmr@gmail.com

Maithri G. S.
Department of Electrical Engineering
Indian Institute of Science
Bengaluru 560012
Email: maithrigs@gmail.com

Sasidharan Ayyavu
Department of Electrical Engineering
Indian Institute of Science
Bengaluru 560012
Email: ayyavusasidharan@gmail.com

Gangotri Nadiger
Department of Electrical Engineering
Indian Institute of Science
Bengaluru 560012
Email: gangotrirg3@gmail.com

***Abstract*—We report our recent work on the recognition of scene text captured by mobile cameras, which we have named Kannada Pado. The text region is currently manually cropped using a user-friendly interface, which permits repeated croppings from the captured image in a hierarchical fashion. The scene text segment is then binarized using the algorithm, midline analysis, and propagation for segmentation. The segmented binary text image is recognized using Lipi Gnani Kannada OCR. The recognized text can be transcribed in Roman, Devanagari, and other principal Indian scripts. Such tools will be of immense use in metropolitan cities such as Bengaluru for business visitors and tourists to be able to read important textual information using their mobile itself. The entire implementation is of low computational complexity and hence, runs fully on the mobile itself, without any backend computation. Currently, text recognition accuracy is the bottleneck, which, when improved, will make the app immediately usable by people. Then, it will be made available to the public from Google Playstore.**

***Index Terms*—KannadaPado, Kannada, scene images, gamma correction, binarization, MAPS, NESP, Android, OCR, transcription, Tamil, Roman, Devanagari, Indic scripts.**

I. Introduction

Bengaluru is a city frequented by a number of business travellers and tourists. However, there are many display boards, road signs and other name boards purely in Kannada, unlike, for example, Singapore, where all the boards have text in the four official languages. Thus, there is a real need for a user-friendly system which helps these visitors make sense out of these Kannada-only display boards.

A mobile-based application that can recognize Kannada text on display boards on shops and offices, sign boards, name boards on buses and trains, menu cards, direction signs, etc. and transcribe the text into a target Indic script of one's choice will be a very useful tool for all business travellers and tourists, who visit Karnataka for a few days. Applications such as Camscanner exist, which work for English. Many text detection algorithms [1]–[3] have been proposed in the literature. EAST text detector [4] by OpenCV is a popular, deep-learning based system for Roman text, which unfortunately does not perform well for scripts such as Kannada.

Further, deep learning based algorithms demand a lot of computation, and where possible, it is preferable to be able to have an algorithm which can be locally computed in the mobile itself. However, developing such a system involves many functional blocks. To begin with, the text regions in the captured scene image must be localized. Then these scene word images need to be extracted and binarized to improve their contrast. These images need to be then recognized by a good Kannada OCR. Finally, we need to have a transliterator or a transcription module in order to convert the recognized text to the script known to the user. Of course, it will be ideal if one can also add the possibility of translating the Kannada text to other languages and reading it!

In this context, we make an attempt to achieve our objective using already published, promising algorithms, which are computationally simple and evaluate the results.

II. Cropping of the Region of Interest

We have created an Android mobile application for recognising Kannada text from camera captured scene images. We have collected about a few hundred images containing

Fig. 1: Sample scene images with Kannada text. The images also show the KannadaPado Android application with the cropping facility enabled. The cropping utility can be invoked multiple times to facilitate selection of the exact region of interest.

TABLE I: Details of the Kannada scene images collected

Nature of image	Number
Road signs	81
Display boards (shops)	108
Display boards (offices)	90
Direction signs	63

Kannada text by photographing sign boards, boards showing destination on buses and trains, display boards on shops and other establishments, road signs, etc. from different locations in Bengaluru and Hoskote. Table 1 shows the distribution of the different categories of the collected images.

Figure 1 shows some sample images from our collection, with the cropping facility enabled. Figure 2 shows the first-level cropped regions from two of the images shown in Fig. 1. As we can see, it is preferable to extract only the text line from the second example image shown. Thus, our app provides the facility to invoke the cropping facility multiple times in a hierarchical fashion, till we are able to exactly get only the required text region. Samples of such text-only regions so cropped are shown in Fig. 3(a).

We have used ScanLibrary (an Android open-source library) for polygonal selection (auto edge detection) for cropping the image and perspective transformation on cropping. This final cropped image is sent to the Kannada OCR-Engine for recognition of the scene text after binarization. We have added transcription feature to view the recognized text in other lndic scripts and share both the images and text to others using the default 'shares options available on

Fig. 2: Two of the sample images shown in Fig.1, after the first level of cropping.

the Android platform (e.g., Mail, WhatsApp, FB, etc.,).

Text recognition from scene word images is difficult due to motion blur, uneven illumination, shadows and defocussing. In order to get better accuracy, we have implemented Midline Analysis and Propagation of Segmentation (MAPS) algorithm [5] for binarizing the image and then use the resultant tiff image in OCR-Engine for recognizing text. OpenCV has been used for image processing in both

ScanLibrary as well as in the above-mentioned algorithm.

A. Binarization using MAPS technique

Binarization is method of converting a document image from a gray scale image consisting of 256 levels into a binary image consisting of 2 levels. Several methods have been reported in the literature to handle the binarization process for scanned document images [6]–[9]. The methods developed have been adapted to suit particular types of variations in the image such as varying illumination, changing contrast and noisy gray values. The images considered by these techniques are scanned gray scale images of documents with large number of pixels, where there is little variation in the context. However, scene images have color pixels, where there is large color and illumination variations between the pixels. Hence, the earlier methods, which were successful on scanned document images could not perform well on scene images. Compared to the normal document images, scene images have very few words. Hence, we need to provide special attention while handling scene images which can improve the detection and the recognition of the words. Even scene word images (obtained by cropping or automated segmentation methods) retain most of the characteristics of scene images such as color values for pixels and drastic illumination changes.

Proper segmentation of the region of interest is the key step that improves the effectiveness of any recognition technology, whether it deals with printed text [10], handwriting [11]–[13], scene images [3], born-digital images [14], texture imaged [15] or medical images [16]. One of the approaches in handling such low quality or low resolution images is to first superresolve the image to enhance its quality [17], [18], before attempting to recognize the same. Special binarization techniques have been proposed [5], [19], [20] to cover the missing link between the scene word images and the regular optical character recognition (OCR) engines. Normally, the variation of illumination is minimal at the word level of a scanned image of printed text. Whereas, this may not be true in scene word images. As explained in the earlier section, each text line is cropped such that the image is suitable for segmentation using MAPS technique.

The recognition engines are developed assuming that any image input to the engine is a gray scale or a binary valued image. The recognition engine do not perform best when the input is not a gray scale or a binary image. Here, MAPS method is used as the link to convert the color word level images into good quality binary images, which are easily handled by the recognition engines.

In MAPS method [5], the color word level image is converted into a gray scale image. The middle line of cropped word image is then selected and pixels that in this line are processed at the first stage. A window size appropriate to the height of the cropped word image is used to counter the illumination variation and varying widths of the characters. The sliding window is used to move across the middle line to compute the mean value of all the pixels falling within the window. If the variation in the pixel values is very less, then the mean value is replaced by the global threshold value which is pre-computed for the middle line. The sliding window is moved from left-to-right and then right-to-left to normalize the mean value from both directions horizontally across the middle line. The mean value is used as a threshold on the middle line pixels. A binarized image for the middle line is obtained after the completion of the first stage.

In the second stage, statistics such as mean and standard deviation for the underlying normal distribution are computed from the the binarized image of the middle line. We get two normal distributions which represent the background and foreground variations. The pixels in the rest of the word image (outside the midline) are classified using Bayesian classifier [21] into background and foreground pixel from the distributions. In some cases, the background and the foreground for the word image would be interchanged. The number of background and foreground pixels along the boundary, at the borders and the maximum width of the background and the foreground components are used to flip the background and the foreground. The results of binarization using MAPS technique are shown in Fig. 3.

B. Text recognition using Lipi Gnani

Several OCRs have been proposed for Kannada in the literature [22], [23]. The binarized images are recognized by Lipi Gnani OCR [10]. This is a state-of-the-art Kannada OCR being marketed by RaGaVeRa Indic Technologies Pvt Ltd., which is a startup incubated by the Society for Innovation and Development, Indian Institute of Science. This OCR has been shown by the authors to perform better than Google's Tesseract version v4.0.0 at the time of publication of the above paper, on a public Benchmark dataset of 251 images taken from old Kannada books.

C. Cross-lingual transcription in Indic scripts

The app is primarily meant for people, who cannot read Kannada. Hence, the output text needs to be rendered in a script known to the user of the app, say Devanagari, Roman or Tamil. Hence the name, KannadaPado.

We make use of the open source MILE transliteration code published by A. G. Ramakrishnan et al. [24] in Github [25]. This makes use of a grapheme-to-phoneme converter for the source language and phoneme sequence is then rendered in the chosen destination script. This was originally developed as a Google Chrome plugin, and has been adapted for the current application.

III. Experiments, Results and Discussion

Experiments are conducted to evaluate the developed system. About ten scene images were used for testing. Figure 3 shows some sample cropped images and their

(a) Cropped scene word images.

(b) Binarized scene word images.

Fig. 3: Examples of cropped images with Kannada text and the corresponding MAPS-binarized images.

binarized counterparts. The quality of the binarized images are reasonably good where the input images have text of the same shade of color. Where the color or its shade changes significantly, the quality of binarization suffers.

The recognition results are not great, with the character level accuracy being around 54%, since we have used an OCR that was developed for good quality, scanned images of printed text. A good dictionary based postprocessing [26] can correct most of the errors since the text we deal with are mostly proper nouns.

A. *Evaluation of the system by prospective users*

The convenience of the user interface and the accuracy of recognition and transcription were tested by a number of prospective users.

IV. Conclusion

We have designed, developed and tested an Android based scene text recognition system for Kannada. The output can be read in the Indic script of choice or Roman by the user.

Handling interspersed English words is the next challenge. To begin with, we would like to identify English words and leave them as image itself after binarization. Further, if the text line is curved, then it is necessary to straighten the text line image, before presenting it to the OCR [27]. Also, we shall be extending the developed system to the sister Dravidian languages, starting with Tamil.

Acknowledgment

The authors thank the Technology Development for Indian Languages, MeitY, Government of India for funding this work under the National Language Translation Mission. Project ref: 11(1)/2020- HCC(TDIL).

References

[1] S Sabari Raju, Peeta Basa Pati, AG Ramakrishnan, "Text localization and extraction from complex color images," Proc. International Symposium on Visual Computing, Springer, Berlin, Heidelberg, pp. 486-493, 2005.

[2] Thotreingam Kasar, Jayant Kumar, AG Ramakrishnan, "Font and background color independent text binarization," Proc. Second international workshop on camera-based document analysis and recognition, 2007.

[3] Thotreingam Kasar, A. G. Ramakrishnan, "Multi-script and multi-oriented text localization from scene images," Proc. International Workshop on Camera-Based Document Analysis and Recognition, 2011.

[4] OpenCV Text Detection (EAST text detector), https://www.pyimagesearch.com/2018/08/20/opencv-text-detection-east-text-detector/ Last accessed on 21 April 2021.

[5] Deepak Kumar, M. N. Anil Prasad and A. G. Ramakrishnan, "MAPS: Midline analysis and propagation of segmentation," Eighth Indian Conf. on Vision, Graphics and Image Processing (ICVGIP), December 16-19, 2012.

[6] N. Otsu, "A Thresholding Selection Method from Gray-level Histogram," IEEE Trans. SMC, vol. 9, pp. 62–66, March 1979.

[7] W. Niblack, "An introduction to digital image processing." New York: Prentice Hall, 1986.

[8] J. J. Sauvola and M. Pietäikinen, "Adaptive document image binarization, Pattern Recognition," vol. 33, no. 2, pp. 225–236, 2000.

[9] D. Bradley and G. Roth, "Adaptive thresholding using the integral image," Journal of Graphics Tools, vol. 12, no. 2, pp. 13–21, 2007.

[10] Shiva Kumar H. R., Ramakrishnan A. G., "Lipi Gnani - A Versatile OCR for Documents in any Language Printed in Kannada Script," ACM Transactions on Asian and Low-Resource Language Information Processing, May 2020.

[11] Suresh Sundaram, A G Ramakrishnan, "Attention-feedback based robust segmentation of online handwritten isolated Tamil words," ACM Transactions on Asian Language Information Processing (TALIP), Vol 12(1), 2013.

[12] Suresh Sundaram and A. G. Ramakrishnan, "A verification-based segmentation of online handwritten words," Indian Patent Office Application No: 3974/CHE/2010.

[13] Suresh Sundaram, A. G. Ramakrishnan, "Lexicon-Free, Novel Segmentation of Online Handwritten Indic Words," Proc. International Conference on Document Analysis and Recognition (ICDAR), pp. 1175-1179, Sept. 2011.

[14] Deepak Kumar and A. G. Ramakrishnan, "OTCYMIST: Otsu-Canny minimal spanning tree for born-digital images," Proc. 10th International Workshop on Document Analysis Systems (DAS 2012), Gold Coast, Queensland, Australia, 2012.

[15] AG Ramakrishnan, S Kumar Raja, HV Raghu Ram, "Neural network-based segmentation of textures using Gabor features," Proc. 12th

IEEE Workshop on Neural Networks for Signal Processing, 2002, pp. 365-374.

[16] Niranjan Joshi, R.Srikanth, A.G.Ramakrishnan, "Model based Bayesian approach for MR Image Segmentation", Proc. V International Conf Adv in Pattern Recognition (ICAPR 2003), ISI, Kolkata, Dec 10-13, 2003.

[17] Ram Krishna Pandey, A G Ramakrishnan, "Efficient Document-Image Super-Resolution Using Convolutional Neural Network," Sadhana, Vol. 43(15), Feb. 2018.

[18] Ram Krishna Pandey, A G Ramakrishnan, "Improving the perceptual quality of document images using deep neural network," Proc. 16th International Symposium on Neural Networks (ISNN 2019), Moscow, Russia, July 10-12, 2019.

[19] Deepak Kumar, A. G. Ramakrishnan, "Power-law transformation for enhanced recognition of born-digital word images," Proc. 2012 International Conference on Signal Processing and Communications (SPCOM), IEEE.

[20] Deepak Kumar, M. N. Anil Prasad and A. G. Ramakrishnan, "NESP: Nonlinear enhancement and selection of plane for optimal segmentation and recognition of scene word images," Document Recognition and Retrieval (DRR) XX, February 5-7, 2013.

[21] R. O. Duda, P. E. Hart and D. G. Stork, "Pattern classification," Wiley, 2001.

[22] B. Vijay Kumar, A. G. Ramakrishnan, "Radial basis function and subspace approach for printed Kannada text recognition," Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, 2004. (ICASSP'04), 2004.

[23] B. Vijay Kumar, A. G. Ramakrishnan, "Machine recognition of printed Kannada text," Proc. Fifth IAPR Workshop on Document Analysis Systems (DAS-02), 2002.

[24] A. G. Ramakrishnan, Royal, Denzil Sequiera, Shashank S Rao, Shiva Kumar H. R., "Transliteration of Indic languages to Kannada with a user-friendly interface," Proc. 2015 IEEE International Advance Computing Conference (IACC).

[25] A. G. Ramakrishnan, Shashank S Rao, "Open source code for MILE-IISc Transliterator," https://github.com/MILE-IISc/MILE-Transliterator. Last accessed 21 April 2021.

[26] Suresh Sundaram, Bhargava Urala, AG Ramakrishnan, "Language models for online handwritten Tamil word recognition," Proc. Workshop on Document Analysis and Recognition (DAR 2012), pp. 42-48, Dec. 2012.

[27] T. Kasar and A. G. Ramakrishnan, "Method of extracting and aligning curvilinear text of a captured image and a system thereof," Patent No. 344198. Indian Patent Office Application No: 109/CHE/2011, 2011.