Electronic Letters on Computer Vision and Image Analysis 13(2):32-34, 2014

Methods for text segmentation from scene images

Deepak Kumar

MILE Laboratory, Department of Electrical Engineering, IISc, Bangalore, INDIA 560012
Advisor/s: A. G. Ramakrishnan
Date and location of PhD thesis defense: 24 February 2014, Indian Institute of Science

Received 19 January 2014; accepted 25 May 2014

1 Abstract

Camera-captured scene/born-digital image analysis helps in the development of vision for robots to read text, transliterate or translate text, navigate and retrieve search results. However, text in such images does not follow any standard layout, and its location within the image is random in nature. In addition, motion blur, non-uniform illumination, skew, occlusion and scale-based degradations increase the complexity in locating and recognizing the text in a scene/born-digital image.

OTCYMIST method [2] is proposed to segment text from the born-digital images. This method won the first place in ICDAR 2011 [9] and placed in the third position in ICDAR 2013 [11] for its performance on the text segmentation task in robust reading competitions for born-digital image data set. Here, Otsu binarization and Canny edge detection are separately carried out on the three colour planes of the image. Connected components (CCs) obtained from the segmented image are pruned based on thresholds applied on their area and aspect ratio. CCs with sufficient edge pixels are retained. The centroids of the individual CCs are used as nodes of a graph. A minimum spanning tree is built using these nodes of the graph. Long edges are broken from the minimum spanning tree of the graph. Pairwise height ratio is used to remove likely non-text components. CCs are grouped based on their proximity in the horizontal direction to generate bounding boxes (BBs) of text strings. Overlapping BBs are removed using an overlap area threshold. Non-overlapping and minimally overlapping BBs are retained for text segmentation. These BBs are split vertically to localize text at the word level.

A word cropped from a document image can easily be recognized using a traditional optical character recognition (OCR) engine. However, recognizing a word, obtained by manually cropping a scene/born-digital image, is not trivial. Existing OCR engines do not handle these kinds of scene word images effectively. Our intention is to first segment the word image and then pass it to the existing OCR engines for recognition. It is advantageous in two aspects: it avoids building a character classifier from scratch and reduces the word recognition task to a word segmentation task. Here, we propose three bottom-up approaches to segment a cropped word image. These approaches choose different features at the initial stage of segmentation.

Power-law transform (PLT) [3] was applied to the pixels of the gray scale born-digital images to non-linearly enhance the histogram. The recognition rate achieved on born-digital word images is 82.9%, which is 20% more than the top performing entry (61.5%) in ICDAR 2011 [9] robust reading competition. The recognition rate is 82.7% and 64.6% for born-digital and scene images of ICDAR 2013 robust reading competition [10], respectively, using PLT.

Correspondence to: <deepak@ee.iisc.ernet.in>

Recommended for acceptance by <Alicia Fornés and Volkmar Frinken>
ELCVIA ISSN:1577-5097
Published by Computer Vision Center / Universitat Autònoma de Barcelona, Barcelona, Spain

In addition, we applied PLT to the colour planes such as red, green, blue, intensity and lightness plane by varying the gamma value. We call this technique as Nonlinear enhancement and selection of plane (NESP) for optimal segmentation [7], which is an improvement over PLT. NESP chooses a particular plane with a proper gamma value based on Fisher discrimination factor. The recognition rate is 72.8% for scene images of ICDAR 2011 robust reading competition, which is 30% higher than the best entry (41.2%). The recognition rate is 81.7% and 65.9% for born-digital and scene images of ICDAR 2013 robust reading competition [10], respectively, using NESP.

Another technique, midline analysis and propagation of segmentation (MAPS) [4], has also been proposed for word segmentation. Here, the middle row pixels of the gray scale image are first segmented and the statistics of the segmented pixels are used to assign text and non-text labels to the rest of the image pixels using min-cut method. Gaussian model is fitted on the middle row segmented pixels before the assignment of other pixels. In MAPS method, we assume the middle row pixels are least affected by any of the degradations. This assumption is validated by the good word recognition rate of 71.7% on ICDAR 2011 robust reading competition for scene images. The recognition rate is 83.8% and 66.0% for born-digital and scene images of ICDAR 2013 robust reading competition [10], respectively, using MAPS. The best reported results for ICDAR 2003 word images is 61.1% using custom lexicons containing the list of test words. On the other hand, NESP and MAPS achieve 66.2% and 64.5% for ICDAR 2003 word images without using any lexicon. By using similar custom lexicon, the recognition rates for ICDAR 2003 word images go up to 74.9% and 74.2% for NESP and MAPS methods, respectively.

We manually segmented word images [1] and recognized these images using OCR to benchmark maximum possible recognition rate for each database [6]. The recognition rates of the proposed methods and the benchmark results are reported on the seven publicly available word image data sets and compared with the results reported in the literature.

We have designed a classifier to recognize Kannada characters and words from Chars74k data set and our own image collection, respectively. Discrete cosine transform (DCT) and block DCT are used as features to train separate classifiers. Kannada words are segmented using the same techniques (MAPS and NESP) and further segmented into groups of components, since a Kannada character may be represented by a single component or a group of components in an image. The recognition rate on Kannada words is reported for different features with and without the use of a lexicon. The obtained recognition performance for Kannada character recognition (11.4%) is three times the best performance (3.5%) reported in the literature [5].

This thesis has dealt with the principal aspects of camera captured scene/born-digital text image analysis: text localization, text segmentation, and word recognition. We have benchmarked the recognition rates of five word image data sets. We conducted a multi-script robust reading competition [8] as part of ICDAR 2013 [11]. This competition was aimed to determine whether the text localization and segmentation methods were capable of handling any text, independent of the script.

References

[1] T. Kasar, D. Kumar, M. N. Anil Prasad, D. Girish and A. G. Ramakrishnan, "MAST: Multi-Script Annotation Toolkit for Scenic Text", *Proc. Joint Workshop on Multilingual OCR and Analytics for Noisy Unstructured Text Data (J-MOCR-AND)*, Sept. 17, 2011, Beijing, China. doi: 10.1145/2034617.2034633

[2] D. Kumar and A. G. Ramakrishnan, "OTCYMIST: OtsuCanny Minimal Spanning Tree for Born-Digital Images", *Proc. 10th IAPR Intl. Workshop on Document Analysis Systems (DAS 2012)*, Queensland, Australia, March 27-29, 2012. doi: 10.1109/DAS.2012.65

[3] D. Kumar and A. G. Ramakrishnan, "Power-law Transformation for Enhanced Recognition of Born-Digital Word Images", *Proc. Intl. Conf. on Signal Processing and Communications (SPCOM)*, July 22-25, 2012, Bangalore. doi: 10.1109/SPCOM.2012.6290009

[4] D. Kumar, M. N. Anil Prasad and A. G. Ramakrishnan, "MAPS: Midline analysis and propagation of segmentation", *Proc. 8th Indian Conf. on Vision, Graphics and Image Processing (ICVGIP)*, December 16-19, 2012. doi: 10.1145/2425333.2425348

[5] D. Kumar and A. G. Ramakrishnan, "Recognition of Kannada characters extracted from scene images", *Proc. Workshop on Document Analysis and Recognition (DAR 2012)*, December 16, 2012. doi: 10.1145/2432553.2432557

[6] D. Kumar, M. N. Anil Prasad and A. G. Ramakrishnan, "Benchmarking recognition results on camera captured word image data sets", *Proc. Workshop on DAR 2012*, December 16, 2012. doi: 10.1145/2432553.2432572

[7] D. 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", *Proc. Document Recognition and Retrieval (DRR) XX*, San Francisco, CA, USA, February 5-7, 2013. doi: 10.1117/12.2008519

[8] D. Kumar, M. N. Anil Prasad and A. G. Ramakrishnan, "Multi-script robust reading competition in ICDAR 2013", *Proc. MOCR 2013*, Washington DC, USA, August 24, 2013. doi: 10.1145/2505377.2505390. http://mile.ee.iisc.ernet.in/mrrc/.

[9] ICDAR 2011 Robust Reading Competition - Challenge 1: Reading Text in Born- Digital Images (Web and Email). *http://www.cv.uab.es/icdar2011competition/.*

[10] ICDAR 2013 Robust Reading Competition. *Proc. 12th ICDAR*, pages 1115-1124, 2013. *http://dag.cvc.uab.es/icdar2013competition/.*

[11] ICDAR 2013 competitions. *http://www.icdar2013.org.*