Abstract

Recognition of text from camera-captured scene/born-digital images help in the development of aids for the blind, unmanned navigation systems and spam filters. However, text in such images is not confined to any page layout, and its location within in 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.

Text localization and segmentation techniques are proposed for the born-digital image data set. The proposed OTCYMIST technique won the first place and placed in the third position for its performance on the text segmentation task in ICDAR 2011 and ICDAR 2013 robust reading competitions for born-digital image data set, respectively. Here, Otsu's binarization and Canny edge detection are separately carried out on the three colour planes of the image. Connected components (CC's) obtained from the segmented image are pruned based on thresholds applied on their area and aspect ratio. CC's with sufficient edge pixels are retained. The centroids of the individual CC's 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. CC's are grouped based on their proximity in the horizontal direction to generate bounding boxes (BB's) of text strings. Overlapping BB's are removed using an overlap area threshold. Non-overlapping and minimally overlapping BB's are used for text segmentation. These BB's 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. In two aspects, it is advantageous: it avoids building a character classifier from scratch and reduces the word recognition task to a word segmentation task. Here, we propose two bottom-up approaches for the task of word segmentation. These approaches choose different features at the initial stage of segmentation.

Power-law transform (PLT) was applied to the pixels of the gray scale born-digital images to non-linearly modify 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 robust reading competition. In addition, we explored applying 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, 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, respectively, using NESP.

Another technique, midline analysis and propagation of segmentation (MAPS), has also been proposed. 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, 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, 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.

In place of passing an image segmented by a method, manually segmented word image is submitted to an OCR engine for benchmarking maximum possible recognition rate for each database. 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 these of reported results in the literature.

Since no good Kannada OCR is available, a classifier is designed 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.