Exploiting Colour Information for Better Scene Text Recognition

Muhammad Fraz[1]
M.Fraz@lboro.ac.uk
M. Saquib Sarfraz[2]
Muhammad.Sarfraz@kit.edu
Eran A. Edirisinghe[1]
E.A.Edirisinghe@lboro.ac.uk

[1] Department of Computer Science
Loughborough University
Loughborough, UK

[2] Computer Vision for Human Computer Interaction Lab
Karlsruhe Institute of Technology
Karlsruhe, Germany

The problem of scene text recognition has gained significant importance because of its numerous applications. A variety of methods has been recently proposed that explore various theoretical and practical aspects to solve this problem. In this work, we focus towards a framework to recognize the text present in outdoor scene images. The text information carries one important property, that is, its colour in comparison to its background. Text information is always placed in such a way that it stands out from its background. In the same way, most of the time the characters in a word possess similar colour that helps us to recognize the letters of a particular word. We exploit this characteristic of text regions to solve the problem of character recognition. The character recognition pipeline is further extended in to a word recognition framework where the estimated word combinations are matched against a lexicon.

The existing approaches for scene text recognition can be roughly divided in to two broad categories: Region grouping based methods and object recognition based methods. In this work, we have combined region grouping method with object recognition based strategy to achieve the advantages of both techniques. First, we binarize the image using colour information and perform foreground segmentation to separate characters from background. Next, we extract shape representation features on binary images and perform character classification using a pre-trained classifier. The recognized characters form words that are fed in to a string similarity matching stage where lexicon based search is performed to find the closest matching word.

Character Identification: We use the bilateral regression [2] for character identification. However, our approach is different than the original method in that we only use it to estimate the horizontal location of each character in word image. The bilateral regression models the foreground pixels by using a weighted regression that assigns weight to each pixel according to its location with respect to foreground in feature space. The pixels that belong to the foreground get high weights in comparison to the pixels belonging to background. In this case, the regression model in equation 1 represents the quadratic surface that best models the image as a function of pixel locations.

$$z = ax^2 + by^2 + cxy + dx + ey + f \tag{1}$$

We enhance the operation of bilateral regression by a pre-processing step where the foreground colour is estimated a priori. We apply n-level colour quantization to achieve binary image for each quantization level. We use Minimum Variance Quantization (MVQ) originally proposed by Heckbert [3]. We quantize each word image in to three colours and analyse the respective binary maps for three quantization levels to estimate the foreground. The characters are cropped from the actual word images using the estimated horizontal location and width from bilateral regression while the height is kept same as the height of the actual word image. the segmented masks are used to crop the characters from original (coloured) image and fed into the character recognition pipeline explained next.

Character Classification: Similar to the character identification stage, we use colour quantization to enhance the character. We found on the basis of extensive experimentation that for a character image 2-level quantization is good enough to recover the full character pixels from background. We therefore generate two binary images corresponding to the two colour levels by assigning the pixels for each colour cluster a value '1' (white similar to the previous stage), we categorize the two binary images as foreground character map by simply analysing the white pixels density along the borders of each binary map. The binary map that possesses the higher total number of corner white pixels is considered as background and the other binary map is classified as the character map. We compute HOG-SVM for character binary map representation and classification.

Word Recognition: The errors in character recognition are inevitable because of high interclass similarity between various characters. In order to find the correct word from various character combinations, the predicted words are aligned with the words available in the lexicon using a string similarity measure. The closest matched word in the lexicon is declared as the word in the image. We adopted a simple strategy where the alignment is performed using Lavenshtien distance.

Figure 1: Improved character identification. Row 1 shows the original images. Row 2 shows the results of character segmentation using Bilateral Regression. Row 3 shows the results of character segmentation using the combination of proposed pre-processing and Bilateral Regression.

Results: The proposed characters recognition pipeline outperforms the current state-of-the-art on Chars74k [1] ICDAR03-CH [5] dataset. Further to that, the proposed word recognition pipeline outperforms the state-of-the-art on challenging ICDAR03-Word [5] and ICDAR11-Word [4] benchmark datasets.

Computational Performance: The proposed framework is implemented in MATLAB. The average execution time for the proposed word recognition pipeline on a standard PC is 1.7 seconds. The separate average execution time for three stages: Character Identification, Character Recognition and Word Recognition is 1.2 sec., 0.4 sec. and 0.1 sec. respectively. Note that the code is unoptimized. The execution time can be further reduced near real-time with the inclusion of code optimization and parallel processing techniques.

Conclusively, the proposed recognition method combines region grouping method with object recognition based strategy to achieve state-of-the-art performance on benchmark datasets. The proposed modification for bilateral regression based segmentation drastically improved character identification performance. The binary maps of the segmented characters have been directly used to extract shape features and fed in to the trained SVM classifier. Finally, a basic string similarity measure has been used to align the estimated words with the lexicon to remove inaccuracies. The experimental results show that proposed framework is accurate, fast, simple and exploitable for practical applications.

[1] T. de. Campos, B. Babu, and M. Verma. Character recognition in natural images. In *VISAPP*, 2009.

[2] Jacqueline L. Feild and Erik G. Learned-Miller. Improving open-vocabulary scene text recognition. In *Proceedings of the 2013 12th International Conference on Document Analysis and Recognition*, ICDAR '13, pages 604–608, Washington, DC, USA, 2013. IEEE Computer Society. ISBN 978-0-7695-4999-6. doi: 10.1109/ICDAR.2013.125.

[3] Paul Heckbert. Color image quantization for frame buffer display. *SIGGRAPH Comput. Graph.*, 16(3):297–307, July 1982. ISSN 0097-8930. doi: 10.1145/965145.801294.

[4] A Shahab, F. Shafait, and A Dengel. Icdar 2011 robust reading competition challenge 2: Reading text in scene images. In *Document Analysis and Recognition (ICDAR), 2011 International Conference on*, pages 1491–1496, Sept 2011. doi: 10.1109/ICDAR.2011.296.

[5] L. P. Sosa, S. M. Lucas, A. Panaretos, L. Sosa, A. Tang, S. Wong, and R. Young. Icdar 2003 robust reading competitions. In *In Proceedings of the Seventh International Conference on Document Analysis and Recognition*, pages 682–687. IEEE Press, 2003.