Text Detection and Recognition in Camera based Images: A Review

Darshan H Y
Department of ISE
DSCE Bangalore, Karnataka
560078

M T GopalKrishna
Department of ISE
DSCE Bangalore, Karnataka
560078

M C Hanumantharaju
Department of ISE
DSCE Bangalore, Karnataka
560078

ABSTRACT

The increase in availability of high performance, low-priced, portable digital imaging devices has created an opportunity for supplementing traditional scanning for document image acquisition. Cameras attached to cellular phones, wearable computers, and standalone image or video devices are highly mobile and easy to use; they can capture images making them much more versatile than desktop scanners. Should gain solutions to the analysis of documents captured with such devices become available, there will clearly be a demand in many domains. Images captured from images can suffer from low resolution, perspective distortion, and blur, as well as a complex layout and interaction of the content and background. In this paper, we present a survey of application domains and technical challenges for the analysis of documents captured by digital cameras. Each method is discussed in brief and then compared against other approaches.

Keywords

Document, Analysis, Processing, Camera-Based Images, Classification.

1. INTRODUCTION

With the increase in growth of camera-based applications available on smart phones and portable devices, understanding the pictures taken by these devices semantically has gained increasing scope from the computer vision community in these years. Among all the information contained in the image, text, which carries semantic information, could provide valuable clues about the content of the image and thus is very important for human as well as computer to understand the scenes. For character recognition in the scene, these methods [4, 13, 19, 18] directly extract features from the original image and uses various classifiers to the character to recognize. While for scene text recognition, since there are no Binarization and Segmentation stages, most methods that exist [19, 18, 11, 10] take up multi-scale sliding window strategy to get the character detection results. As sliding window strategy does not make use of the special structure information of each character, it will produce many false positives. Thus, these methods mainly depends on the post processing methods such as pictorial structures [19, 18]. As proved by Judd et al. [7], given an image containing text and other objects, viewers tend to fixate on text, suggesting the importance of text to human. In fact, text recognition is indispensable for a lot of applications such as automatic sign reading, language translation, navigation and so on.

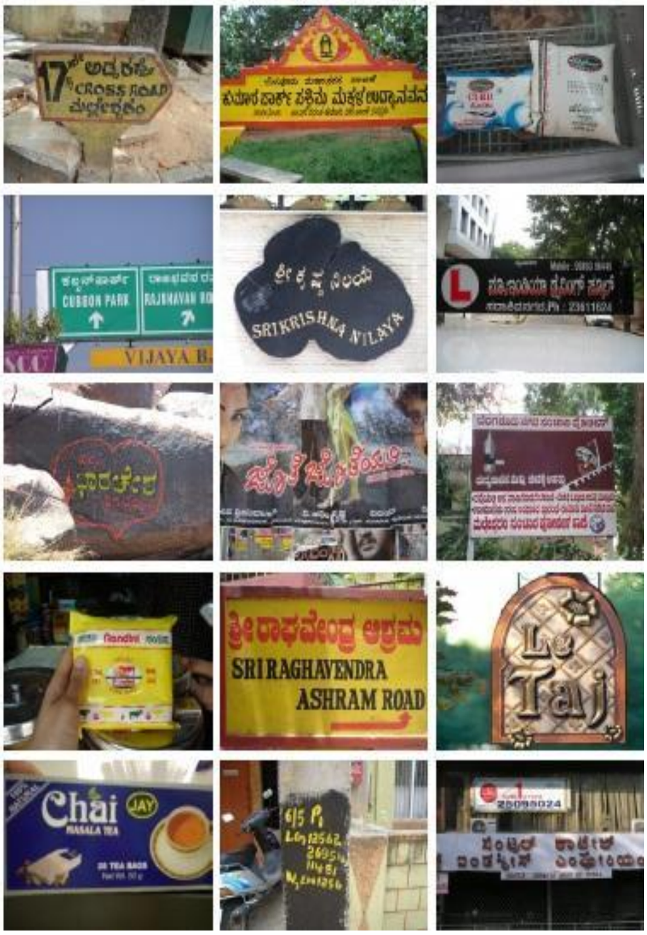

Fig. 1.Sample source images in our data set.

Thus, understanding scene text ismore important than ever. The following sources of variability still need to be accounted for: (a) font style and thickness; (b) background as well as foreground color and texture; (c) camera position which can introduce geometric distortions; (d) illumination and (e) image resolution. All these factors combine to give the problem a flavor of object recognition.

2. DETECTION AND RECOGNITION METHODS

Cunzhao Shi et al. [1] proposed a novel scene text recognition method using part-based tree-structured character detection, different from conventional multi-scale sliding window character detection strategy, which does not make use of the character-specific structure information. They have used part-based tree-structure to model each type of character so as to detect and recognize the characters at the same time. However, Since the text in natural images differs from text in traditional scanned document in terms of resolution, illumination condition, size and font style, the Binarization result is usually unsatisfactory. Moreover, the loss of information during the Binarization process is almost unrecoverable, which means if the Binarization result is poor, the chance of correctly recognizing the text is quite small. As shown in Figure 2, the Binarization result is very disappointing, making it almost impossible for the following Segmentation And the Recognition. We build a tree-structure based model for each type of character. Figure 3 shows the models for some characters. Each rectangle corresponds to a part-based filter of the character and the red lines illustrate the topological relations of the parts. Examples of word recognition results of the proposed method as shown in Figure 4.

Devvrat C. Nigam et al. [2] proposed character extraction and edge detection algorithm for training the neural network to classify and recognize the handwritten characters. In general, handwriting recognition is classified into two types as off-line and on-line handwriting recognition methods. The on-line Methods have been shown to be superior to their off-line counterparts in recognizing handwritten characters due to the temporal information available with the former. There are basically two main phases in our Paper: Pre-processing and Character Recognition. In the first phase, they are preprocessing the given scanned document for separating the Characters from it and normalizing each characters.

Initially we specify an input image file, which is opened for reading and preprocessing. The image would be in RGB format (usually), so we convert it into binary format. To do this, it converts the input image to grayscale format (if it is not already an intensity image), and then uses threshold to convert this grayscale image to binary i.e. all the pixels above a certain threshold as 1 and below it as 0.The model proved here is as in Figure 5.ShangxuanTian et al. [3] propose to recognize the scene text by using an extension of the HOG, namely, co-occurrence HOG (Co- HOG) [13] that captures gradient orientation of neighboring pixel pairs instead of a single image pixel. Co-HOG divides the image into blocks with no overlap which is more efficient than HOG with overlapped blocks [25]. This is essential in the real-time text recognition system. More Importantly, relative location and orientation are considered with each neighboring pixel, respectively, which is moreprecise to describe the character shape.In addition, Co-HOG Keeps the advantages of HOG, i.e., the robustness to varying illumination and local geometric transformations.

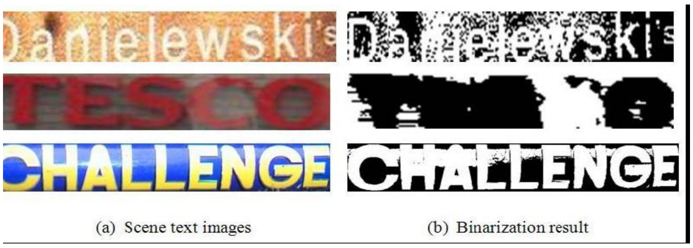

Fig. 2. Some scene text binarization results

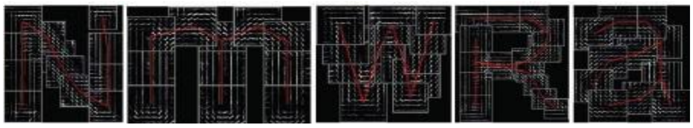

Fig. 3. Tree-structured models for some characters

Jerod J. Weinman et al. [5] propose a probabilistic graphical model for STR that brings both bottom-up and top-down information as well local and long-distance relationships into a single elegant framework. In addition to individual character appearance, our model integrates appearance similarity, one underused source of information, with local language statistics and a lexicon in a unified probabilistic framework to reduce false matches errors in which the different characters are given the same label by a factor of four and improve overall accuracy by greatly reducing word error. The model adapts to the data present in a small sample of text, as typically encountered when reading signs, while also using higher level knowledge to increase robustness.

MohanadAlata et al. [6] proposed method was based on a combination of an Adaptive Color Reduction (ACR) technique and a Page Layout Analysis (PLA) approach. K. AtulNegi, Nikhil Shanker and Chandra KanthChereddi [6] presented a system to locate, extract and recognize Telugu text. The circular nature of Telugu script was exploited for segmenting text regions using the Hough Transform.

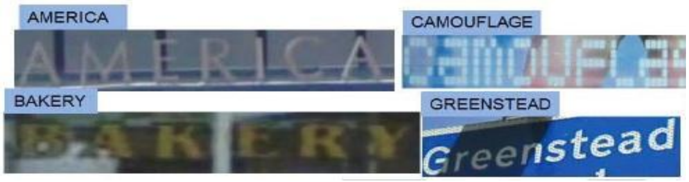

Fig. 4.Examples of word recognition results of the proposed methods

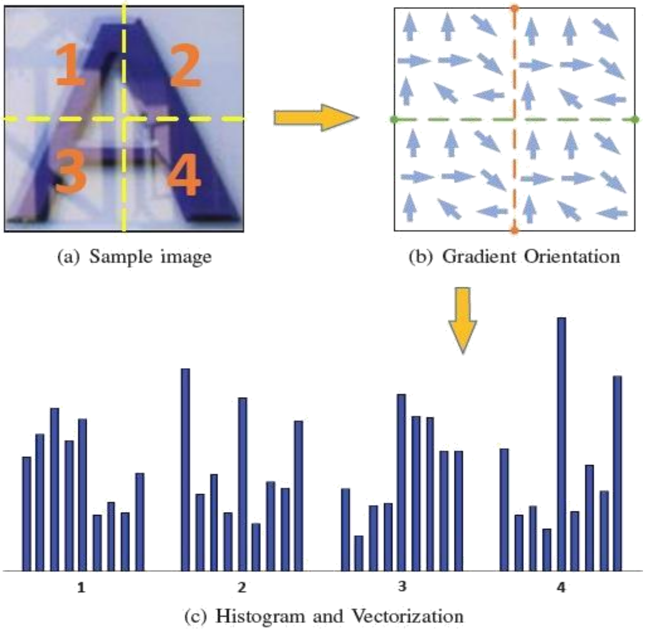

Fig. 5. Illustration of HOG feature extraction: (a) shows a character sample which is divided into 4 blocks (the blocks overlay with neighboring blocks in implementation). (b) shows the corresponding gradient orientation of each block. (c) shows the histogram of gradient orientation and concatenated one after another to form a HOG feature vector normalized.

Isgro et al. [8] have proposed a feature based image mosaicking method. In this method, feature points are firstly extracted from one of two images to be stitched, and then the corresponding points in the other image are calculated. From the corresponding points, a Euclidean transformation parameter between the two images is estimated. The images are stitched using the Euclidean transformation parameter. In the method, an efficient processing is realized by step-by-step matching of feature points. However, the method is also unable to deal with scaling and perspective distortion because the Euclidean transformation includes only translation and rotation. Moreover, it cannot deal with significant rotation since the correlation technique is used for finding the corresponding points. Therefore, the method cannot likewise realize mosaicking of camera captured document images.

Jain et al. [14] perform a color space reduction followed by color segmentation and spatial regrouping to detect text. Although processing of touching characters is considered by the authors, the segmentation phase presents major problems in the case of low quality documents, especially video sequences.

The video indexing system introduced by Sato et al. combines closed caption extraction with superimposedcaption (artificial text) extraction. The text extraction algorithm is based on the fact that text consists of strokes with high contrast and it searches for vertical edges which are grouped into rectangles. The authors recognized the necessity to improve the quality of the text before passing an OCR step. Consequently, they perform an interpolation of the detected text rectangles before integrating multiple frames into a single enhanced image by taking the minimum/maximum value for each pixel. They also introduced an OCR step based on a correlation measure. Most of the work in this field is based on locating and rectifying the text areas (e.g. (Kumar et al., 2007), (Kemp et al., 2002), (Clark and Mirmehdi, 2002) and (Brown et al., 2007)), followed by the application of OCR techniques (Kise and Doermann, 2007). Such approaches are therefore limited to scenarios where OCR works well. For natural scenes, some researchers have designed systems that integrate text detection, segmentation and recognition in A single framework to accommodate contextual relationships. For instance, (Tu et al., 2005) used insights from natural language processing and present a Markov chain framework for parsing images. (Jin and Geman, 2006) Introduced composition machines for constructing probabilistic hierarchical image models which accommodate contextual relationships. This approach allows re-usability of parts among multiple entities and non-Markovian distributions. (Weinman and Learned Miller, 2006) Proposed a method that fuses image features and language information (such as bi-grams and letter case) in a single model and integrates dissimilarity information between character images.

3. CONCLUSION

In this paper we discussed few of the methods aimed at Document Image Processing. We presented a comprehensive review of Detection and recognition methods. From the review, it is obvious that the results produced from the different techniques are best suited for detecting and recognition of text in camera based images. This idea cannot be turn into product that easily unless more effort and improvement are done. There are still a lot of rooms for improvement for this project and also the idea is still in its infancy stage. I would like to suggest that the more research and more improvements are done to it before put it to test on the market. We are taking up the work of images with multi-oriented scene texts, with a more challenging dataset that is NEOCR(Natural Environmental OCR).

4. REFERENCES

[1] K. Wang, B. Babenko, and S. Belongie. End-to-end scene text recognition. In International Conference on Computer Vision (ICCV), 2011.

[2] Shahab, A., Shafait, F., Dengel, A.: Icdar 2011 robust reading competition challenge 2: Reading text in scene images. In: International Conference on Document Analysis and Recognition. (2011).

[3] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained partbased models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(9):16271645, 2010

[4] A. Shahab, F. Shafait, and A. Dengel. Icdar 2011 robust reading competition challenge 2: Reading text in scene images. In International Conference on Document Analysis and Recognition (ICDAR), pages 14911496. IEEE, 2011

[5] NEGI, A.-SHANKER, K. N.-CHEREDDI, C. K. : Localization, Extraction and Recognition of Text in Telugu Document Image, ICDAR03.2003, 1193-1197

[6] Y. Boureau, F. Bach, Y. LeCun, and J. Ponce, Learning mid-level features for recognition, in Computer Vision and Pattern Recognition, 2010

[7] J. J. Weinman, Typographical features for scene text recognition, in Proc. IAPR International Conference on Pattern Recognition, Aug. 2010, pp. 39873990.

[8] N. Sharma, U. Pal, F. Kimura, ”Recognition of Handwritten Kannada Numerals”, 9th International Conference on Information Technology (ICIT’06), ICIT, pp. 133-136.

[9] Rafael C. Gonzalez, Richard E. woods and Steven L.Eddins, Digital Image Processing using MATLAB, Pearson Education, Dorling Kindersley, South Asia,

[10] A. Mishra, K. Alahari, and C. V. Jawahar. Top-down and bottom-up cues for scene text recognition. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition(CVPR),2012

[11] V.K. Govindan and A.P. Shivaprasad, Character Recognition A review, Pattern Recognition, vol. 23, no. 7, pp. 671- 683,.

[12] A. Newell and L. Griffin. Multiscale histogram of oriented gradient descriptors for robust character recognition. In International Conference on Document Analysis and Recognition (ICDAR), pages 1085-1089. IEEE, 2011.

[13] A.K. Jain and B. Yu. Automatic Text Location in Images and Video Frames. Pattern Recognition, 31(12):2055-2076, 1999

[14] D. C. Ciresan, U. Meier, J. Masci, L. M. Gambardella, and J. Schmidhuber. High performance neural networks for visual object classification. Technical Report IDSIA-01-11, DalleMolle Institute for Artificial Intelligence, 2011

[15] L. Neumann and J. Matas. A method for text localization and recognition in real-world images. In ACCV, 2010

[16] A. Mishra, K. Alahari, and C. V. Jawahar. Top-down and bottom-up cues for scene text recognition. In CVPR, 2012

[17] K. Wang, B. Babenko, and S. Belongie. End-to-end scene text recognition. In International Conference on Computer Vision (ICCV), 2011.

[18] K. Wang and S. Belongie. Word spotting in the wild. Computer Vision-ECCV, pages 591-604, 2010

[19] Scene Text Recognition using Part-based Tree-structured Character Detection,2013, Cunzhao Shi, Chunheng Wang,

[20] Baihua Xiao, Yang Zhang, Song Gao and Zhong Zhang State Key Laboratory of Management and Control for Complex Systems, CASIA, Beijing, China,IEEE

[21] Scene Text Recognition using Co-occurrence of Histogram of Oriented Gradients,2013, ShangxuanTian, Shijian Lu y, Bolan Su and Chew Lim Tan.TEXT DETECTION AND CHARACTER RECOGNITION USING FUZZY IMAGE PROCESSING,2012MohanadAlata Mohammad Al-Shabi

[22] Scene Text Recognition using Similarity and a Lexicon with Sparse Belief Propagation,2012 Jerod J. Weinman, Member, IEEE, Erik Learned-Miller, Member, IEEE, Allen R. Hanson Member, IEEE

[23] Teofilo E. de Campos,BodlaRakeshBabu,Manik Varma:2013,CHARACTER RECOGNITION IN NATURAL IMAGES

[24] T. Watanabe, S. Ito, and K. Yokoi, Co-occurrence histograms of oriented gradients for human detection, Information and Media Technologies, vol. 5, no. 2, pp. 659667, 2010

[25] Character Recognition Using Matlabs Neural Network Toolbox:2013, Kauleshwar Prasad, Devvrat C. Nigam, AshmikaLakhotiya and DheerenUmre