Text String Recognition from Natural Scenes by Character Descriptor Map and Stroke Width

Pallavi Suresh Umap
Computer Engineering department
DYPSOET, Pune

Sathishkumar Penchala
Computer Engineering department
DYPSOET, Pune

ABSTRACT
Text detection and recognition in natural scene can give valuable information for many applications. However, getting text from images with complex background is challenging task due to less frequency of occurrence text and presence of background outliers resembling text characters. In text detection, algorithms from previous work are applied to localize text region in scene image. First, character descriptor is employed to extract structure features. Second, we tend to designed novel feature representation, stroke configuration map using character boundary and skeleton to build character structure. Our algorithm style is improved to compatible with mobile application. Developed algorithm style is compatible with the appliance of scene text extraction in good mobile devices. The Android-based demo system is developed to highlight the effectiveness of the method of scene text extraction from nearby objects. Also demo system gives the detailed information about algorithm design and performance improvement of text extraction from natural image. The demo system conjointly provides United States some insight into rule design and performance improvement of scene text extraction. The analysis results on benchmark knowledge sets demonstrate that our projected theme of text recognition is comparable the best existing ways. The evaluation results on benchmark datasets demonstrate that proposed framework outperforms in comparison to best existing ways.

Keywords
Scene text detection and recognition, character descriptor, stroke configuration, text understanding, text retrieval, mobile application, text localization

1. INTRODUCTION
Image-based text information serves as an important indicator in many applications. It provides instructions and presentations for navigation, assistive reading, geocoding, and content-based image retrieval etc. In natural scene images and videos, text characters and strings usually appear in nearby sign boards and hand-held objects. Text-based tags are much more applicable than barcode or fast response code [1] as a result of the latter techniques contain restricted data and require pre-installed marks. Scene text detection and recognition algorithms are necessary to extract text information from scene image by mobile devices. Extracting text from natural scene images must solve two challenging problems. (1) Littered backgrounds with noise and non-text outliers. (2) Numerous text patterns like character varieties, fonts, and sizes. The frequency of prevalence of text in natural scene is incredibly low, and a restricted range of text characters are embedded into advanced non-text background outliers. Background textures, such as grid, window, and brick, even gibe text characters and strings. Though these difficult factors exist in face and automotive, several progressive algorithms [2], [7] have incontestable effectiveness on those applications, because face and automotive, have comparatively stable options. As an example, a frontal face commonly contains a mouth, a nose, two eyes, and 2 brows as previous data. However, it's tough model the structure of text characters in scene pictures due to the shortage of discriminative pixel-level look and structure options from non-text background outliers. Further, text consists of various words wherever every word could contain different characters in varied fonts, styles, and sizes, resulting in massive intra-variations of text patterns.

However, it is difficult to model character structure due to lack of discriminative pixel level appearance and structure features from scene image in presence of variant background interferences .Further, text consists of different words where each word may contain different characters in various fonts, styles, and sizes, resulting in large intra-variations of text patterns. To solve these problems, scene text extraction is separated into two processes [10]: text detection and text recognition. Text detection is used to localize image regions containing text characters and strings. Text recognition is used to convert pixel based text into readable code. It also aims to distinguish different text characters and properly composed text words. Text information in natural scene images can give important indicator in many image-based applications. Detection of text and classification of characters in scene images is a challenging visual recognition difficulty for visually challenged people. Most optical character recognition (OCR) systems achieve perfect recognition rate on printed text in scanned documents, but perform poorly when text is embedded into complex background because of background interferences and low frequency of occurrence of text. Thus we need to detect image regions containing text strings and their corresponding orientations. This is compatible with the detection and localization procedure described in the survey of text extraction algorithms .In scene text detection process, we apply the methods presented in our previously proposed paper by C. Yi and Y. Tian [9]. Color-based partition and text line grouping (CT) are performed to extract text strings with arbitrary orientations. Text strings in natural scene images usually appear in alignment, namely, each text character in a text string must possess character siblings at adjacent positions. The structure features among sibling characters can be used to determine whether the connected components belong to text characters or unexpected noises.

2. LITERATURE SURVEY
N. Dalal and B. Triggs detector [5], used a single filter on histogram of oriented gradients (HOG) features for representing category of object. This detector uses a sliding window approach, where a filter is applied at all positions and scales of an image. We can think of the detector as a classifier which taking image, position within that image, and a scale as input. The classifier determines whether or not there is an

instance of the target category at the given position and scale. Major innovation of detector was the construction of particularly effective features. Weinman and Learned-Miller [4] showed that to improve scene text recognition the similarity among characters, in addition to the appearance of the characters with respect to a model, could be used. Here, the Gabor-based appearance model and a language model related are combined to simultaneity frequency and letter case, similarity model, and lexicon model to perform scene character recognition. In the Robust Reading Competition of International Conference on Document Analysis and Recognition (ICDAR) 2011 [3], the best word recognition rate for scene images was only about 41.2%. Reading text form scene image consisted of two tasks namely (1) text localization task and (2) word recognition task. The purpose of text localization task is to identify text regions in scene images and mark their location with axis-aligned rectangular bounding boxes. In word recognition task, cropped word images of scene text are recognized. Cropping was done based on ground-truth word bounding boxes to evaluate recognition performance independently from text localization accuracy. Task of page text localization can be evaluated using any standard methodology for evaluating page segmentation performance that takes into account different categories of segmentation errors. In evaluation of accuracy of word recognition, the edit distance is used with equal cost of deletions, substitutions, and insertions. We normalize the edit distance by the number of characters in ground truth word. Edge pixels at boundaries are obtained from either large neighboring color differences that are greater than a threshold of Canny detector or the 8-neighborhood connection to an existing edge pixel. Their color values are used as observation of the color-pair across two sides of the boundary where the edge pixel is located. We denote color with lower intensity component by ClL and the other one by ClH. RGB space, color values ClL and both have three dimensions. If the boundary belongs to a text character or string, the ClL and ClH represent colors of text and attachment surface respectively. the coordinates of SPx and SPy the edge pixel Pe are used as observation of spatial positions. Then an observation vector x of the edge pixel can be defined by cascading the color values of the two pixels with maximum color difference in neighborhood and the spatial coordinates of the central edge pixel. Thus edge pixel Pe is described by an observation vector x=[ClL , ClH , SPx , SPy] which is a 12-dimesnional point in observation space. To extract text boundaries from scene images, we cluster the observation points of edge pixels into several groups such that edge pixels with similar color-pairs and spatial positions are assigned into identical boundary layer. In this process, GMM is employed to analyze the distributions of observation points of edge pixels. At first k-means clustering is applied to calculate k centers of observation points, which are used as initial means μi $(1 \leq i \leq k)$ of the Gaussian mixture distributions. Then the corresponding k variances σi $(1 \leq i \leq k)$ are calculated from the means of observation points. Thus we can initialize a group of Gaussian distributions Next, over the observation points of edge pixels, EM algorithm is applied to obtain maximum likelihood estimate of the GMM parameters, including weights, means, and variances of the k Gaussian distributions. In EM process, the GMM parameters are iteratively updated from their initial values derived by k-means clustering. Then boundary layer is built from each of the Gaussian distributions under the parameters derived by EM. For an edge pixel, if it generates maximum likelihood in the i-th Gaussian distribution, it will be assigned into the i-th boundary layer Bi [8].

3. PROPOSED SYSTEM

The scene text recognition method combines scene text detection and scene text recognition algorithms. In text localization text regions in scene images are identified and their location is marked with axis-aligned rectangular bounding boxes. From captured image, textual regions within image are localized. Localizing text in scene image is very expensive task. As focus is on independent analysis of single character, text string structure is more robust to distinguish background interferences from text information. [9] The system only considers textual regions and complex backgrounds are not within scope of project. Core functionality of word recognition task is to recognize cropped word images of scene text. Cropping was done based on ground-truth word bounding boxes to evaluate recognition performance independently from text localization accuracy.

By the character recognizer, text understanding is able to provide surrounding text information for mobile applications, and by the character classifier of each character class, text retrieval is able to help search for expect objects from environment. Similar to other methods, the state of-the-art low-level feature descriptors and coding/pooling schemes are used to demonstrate effectiveness of proposed system. Different from other methods, proposed method combines the low-level feature descriptors with stroke configuration to model text character structure. Also, we present the respective concepts of text understanding and text retrieval and evaluate our proposed character feature representation based on the two schemes in our experiments. Besides, previous work rarely presents the mobile implementation of scene text extraction, and we transplant our method into an Android-based platform.

3.1 System Architecture

First, letters are grouped into text lines. As single letter usually do not appear in an image, finding groups of letters allows us to remove randomly scattered noise. At the next step of the algorithm, the candidate pairs determined above are clustered together into chains. Initially, each chain consists of a single pair of letter candidates. Two chains can be merged together if they share one end and have similar direction. The process ends when no chains can be merged. Finally, text lines are broken into separate words, using a heuristic that computes a histogram of horizontal distances between consecutive letters [7].

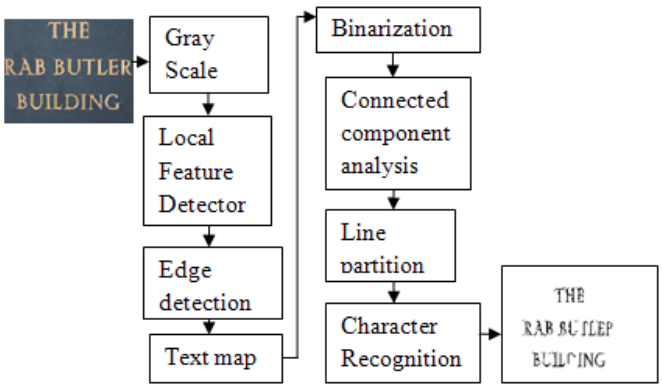

Fig 1.Flowchart of Proposed Framework of Scene Text Recognition

Figure 1, demonstrates that different steps included in text recognition process. SIFT is employed as local feature descriptor and Canny Edge detector is used for edge detection. SIFT algorithm takes an image and transform it into collection

of local feature vectors. Gray scale removes all color information. In the process of edge detection noise is removed. Text mapping gives recognized text in one color and background in different color. In binarization step darkness of color is removed and detected edges are placed on black window.

Binarization is used to get the binary image which is directly fed into Connected Component (CC) analysis. From the binarized text image, we group the connected components (CCs) into text lines. CC analysis is used to find individual characters by grouping pixels into regions using connected component analysis assuming that pixels belonging to the same character have similar properties.

3.2 Problem Definition

The color uniformity and horizontal alignment were employed to localize text regions in scene images. We integrate the functional modules of scene text detection and text recognition. It is able to detect regions of text strings from complex background, and recognize characters in the text regions.

3.3 Mathematical Model

Let V be the variation and a be the aspect ratio of a MSER, the aspect ratios of characters are expected to fall in [a_{min}, a_{max}], the regularized variation V is defined as function of Θ1 and Θ2. Where Θ1 and Θ2 are penalty parameters. Based on experiments on the training database, these parameters are set as Θ1= 0.01, Θ2 = 0.35, a_{max} = 1.2 and a_{min}= 0.3.We use the weighted sum of features as the distance function. Given two data points u, v, let $x_{u,v}$ be the feature vector characterizing the similarity between u and v, and the distance between u and v is defined as $d(u,v;w)=w^T x_{u;v}$. Where w, the feature weight vector together with the threshold T, can be learned using the proposed distance metric learning algorithm.

3.4 Limitations of the Proposed Framework

(1) The proposed framework is based on color uniformity and horizontal alignment of text strings with more than 2 characters, so it cannot handle a text string with non-uniform colors, single character, or text string whose angle with the horizontal is larger than 20 degrees.

(2) In addition, the framework requires enough resolution of the text to be localized. The characters and strings cannot be too small or too blurred. Words with spaces cannot be recognized.

4. CONCLUSION AND FUTURE SCOPE

In this paper a method to localize text regions under complex background and multiple text patterns is designed. It detects text regions from natural scene image/video, and recognizes text information from the detected text regions. In scene text detection, layout analysis of color decomposition and horizontal alignment is performed to search for image regions of text strings. In scene text recognition, two schemes, text understanding and text retrieval, are respectively proposed to extract text information from surrounding environment. Our proposed character descriptor is effective to extract representative and discriminative text features for both recognition schemes. To model text character structure for text retrieval scheme, we have designed a novel feature representation, stroke configuration map, based on boundary and skeleton. Quantitative experimental results demonstrate that our system outperforms over existing text recognition systems.

The low width part of word from complex background cannot be removed as a part of noise. An accuracy rate of text detection should be improved.

5. ACKNOWLEGMENT

This is a great pleasure & immense satisfaction to express my deepest sense of gratitude & thanks to everyone who has directly or indirectly helped me in completing my work successfully express my gratitude towards project guide Prof. Sathishkumar Penchala Department of Computer Engineering. D.Y. Patil School of Engineering and Technology, Pune Dist. Pune Maharashtra, India, who guided & encouraged me in completing this work in scheduled time. I would like to thanks our Principal Dr. Uttam Kalawane, for allowing me to pursue my project in this institute.

6. REFERENCES

[1] Y. Liu, J. Yang, and M. Liu, "Recognition of QR code with mobile phones," in *Proc. CCDC*, Jul. 2008, pp. 203–206.

[2] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection", IEEE Conf. Comput. Vis. Pattern Recognit., 886–893, Jun. 2005.

[3] A. Shahab, F. Shafait, and A. Dengel, "ICDAR 2011 robust reading competition challenge 2: Reading text in scene images," in Proc. Int. Conf. Document Anal. Recognit., Sep. 2011, pp. 1491–1496.

[4] J. J. Weinman, E. Learned-Miller, and A. R. Hanson, "Scene text recognition using similarity and a lexicon with sparse belief propagation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 10, pp. 1733–1746, Oct. 2009.

[5] Y. Liu, J. Yang, and M. Liu, "Recognition of QR code with mobile phones", CCDC, 203–206, July 2008.

[6] E. Ohbuchi, H. Hanaizumi, and L. A. Hock, "Barcode readers using the camera device in mobile phones", Int. Conf. Cyberworlds, 260–265, November 2004.

[7] P. Viola and M. J. Jones, "Robust real-time face detection", Int. J.Comput.Vis., vol. 57, no.2,137–154, 2004.

[8] C. Yi and Y. Tian, "Localizing text in scene images by boundary clustering, stroke segmentation, and string fragment classification," IEEE Trans. Image Process., vol. 21, no. 9, pp. 4256–4268, Sep. 2012.

[9] C. Yi and Y. Tian, "Text string detection from natural scenes by structure-based partition and grouping," IEEE Trans. Image Process.,vol. 20, no. 9, pp. 2594–2605, Sep. 2011.

[10] J. Zhang and R. Kasturi, "Extraction of text objects in video documents: Recent progress," in Proc. 8th IAPR Int. Workshop DAS, Sep. 2008, pp. 5–17.