International Journal of Emerging Engineering Research and Technology
Volume 3, Issue 11, November 2015, PP 152-156
ISSN 2349-4395 (Print) & ISSN 2349-4409 (Online)

Text Detection from Natural Image using MSER and BOW

[1]K.Sowndarya Lahari, [2]M.Haritha, [3]P.Prasanna Murali Krishna

[1] (M.Tech), DECS, DR.Sgit, Markapur, India.
[2]Associate Professor, Department of ECE, DR.Sgit, Markapur, India.
.[3]H.O.D Department of ECE, DR.Sgit, Markapur, India.

ABSTRACT

Text characters and strings in natural scene can provide valuable information for many applications. Extracting text directly from natural scene images or videos is a challenging task because of diverse text patterns and variant background interferences. This project proposes a method of scene text recognition from detected text regions. In text detection, our previously proposed algorithms are applied to obtain text regions from scene image. First, we design a discriminative character descriptor by combining several state-of-the-art feature detectors and descriptors. Second, we model character structure at each character class by designing stroke configuration maps. Our algorithm design is compatible with the application of scene text extraction in smart mobile devices. An Android-based demo system is developed to show the effectiveness of our proposed method on scene text information extraction from nearby objects. The demo system also provides us some insight into algorithm design and performance improvement of scene text extraction. The evaluation results on benchmark data sets demonstrate that our proposed scheme of text recognition is comparable with the best existing methods.

INTRODUCTION

Recognition of text in natural scene images is becoming a prominent research area due to the widespread availability of imaging devices in low-cost consumer products like mobile phones. Detecting text in natural images, as opposed to scans of printed pages, faxes and business cards, is an important step for a number of Computer Vision applications, such as computerized aid for visually impaired and robotic navigation in urban environments. Retrieving texts in both indoor and outdoor environments provides contextual clues for a wide variety of vision tasks. The increasing market of cheap cameras, natural scene text has to be handled in an efficient way.

Text in the image contains useful information which helps to acquire the overall idea behind the image. Character extraction from image is important in many applications. It is a difficult task due to variations in character fonts, sizes, styles and text directions, and presence of complex backgrounds and variable light conditions. Several methods for text (or character) extraction from natural scenes have been proposed. If we develop a method that extracts and recognizes those texts accurately in real time, then it can be applied to many important applications like document analysis, vehicle license plate extraction, text- based image indexing, etc and many applications have become realities in recent years

Some works deal with text detection in the image while more recent ones point out the challenge of text extraction and recognition. Reading text from photographs is a challenging problem that has received a significant amount of attention. Two key components of most systems are (i) text detection from images and (ii) character recognition, and many recent methods have been proposed to design better feature representations and models for both.

LITERATURE SURVEY

Skeleton Pruning by Contour Partitioning with Discrete Curve Evolution (2007), Author Names: X. Bai, L. J. Latecki, And W.-Y. Liu.

In this framework we, introduce a new skeleton pruning method based on contour partitioning. Any contour partition can be used, but the partitions obtained by Discrete Curve Evolution (DCE) yield

Address for correspondence:

excellent results. The theoretical properties and the experiments presented demonstrate that obtained skeletons are in accord with human visual perception and stable, even in the presence of significant noise and shape variations, and have the same topology as the original skeletons. In particular, we have proven that the proposed approach never produces spurious branches, which are common when using the known skeleton pruning methods. Moreover, the proposed pruning method does not displace the skeleton points. Consequently, all skeleton points are centers of maximal disks. Again, many existing methods displace skeleton points in order to produces pruned skeletons.

Advantage

- Discrete Curve Evolution (DCE) excellent results in accord with human visual perception and stable.

Disadvantage

- proposed pruning method does not displace the skeleton points

A Weighted Finite-State Framework for Correcting Errors in Natural Scene OCR (2007), Author Names: R. Beaufort and C. Mancas-Thillou.

The increasing market of cheap cameras, natural scene text has to be handled in an efficient way. Some works deal with text detection in the image while more recent ones point out the challenge of text extraction and recognition. We propose here an OCR correction system to handle traditional issues of recognizer errors but also the ones due to natural scene images, i.e. cut characters, artistic display, incomplete sentences (present in advertisements) and out- of-vocabulary (OOV) words such as acronyms and so on. The main algorithm bases on finite-state machines (FSMs) to deal with learned OCR confusions, capital/accented letters and lexicon look-up. Moreover, as OCR is not considered as a black box, several outputs are taken into account to inter mingle recognition and correction steps. Based on a public database of natural scene words, detailed results are also presented.

Advantage

- The proposed system can handle traditional issues of recognizer errors but also the ones due to natural scene images.

Disadvantage

- The system is time consuming because is not considered as a black box, several outputs are taken into account to intermingle recognition and correction steps.

Automatic Detection and Recognition of Signs from Natural Scenes Author Names: R. Beaufort and C. Mancas-Thillou.

In this paper, we present an approach to automatic detection and recognition of signs from natural scenes, and its application to a sign translation task. The proposed approach embeds multi resolution and multi scale edge detection, adaptive searching, color analysis, and affine rectification in a hierarchical framework for sign detection, with different emphases at each phase to handle the text in different sizes, orientations, color distributions and backgrounds.

Advantages

- The proposed system can handled the multi resolution and multi scale edge detection, adaptive searching, color analysis, and affine rectification successfully.

Disadvantages

- The system cannot restore the image details.

Text Detection and Character Recognition in Scene Images with Unsupervised Feature Learning, Author Names: A. Coates et al.

We use affine rectification to recover deformation of the text regions caused by an inappropriate camera view angle. The procedure can significantly improve text detection rate and optical character recognition (OCR) accuracy. Instead of using binary information for OCR, we extract features from an intensity image directly. We propose a local intensity normalization method to effectively handle lighting variations, followed by a Gabor transform to obtain local features, and finally a linear discriminant analysis (LDA) method for feature selection. We have applied the approach in developing a Chinese sign translation system, which can automatically detect and recognize Chinese signs as input from a camera, and translate the recognized text into English.

Reading text from photographs is a challenging problem that has received a significant amount of attention. Two key components of most systems are (i) text detection from images and (ii) character recognition, and many recent methods have been proposed to design better feature representations and models for both. In this paper, we apply methods recently developed in machine learning -- specifically, large-scale algorithms for learning the features automatically from unlabeled data -- and show that they allow us to construct highly effective classifiers for both detection and recognition to be used in a high accuracy end-to-end system.

Advantage

- The proposed method can automatically detect and recognize Chinese signs as input from a camera, and translate the recognized text into English.

Disadvantage

- The machine learning process need to be maintain more data base which leads to time consuming .

PROPOSED SYSTEM

Layout-Based Scene Text Detection

Here we present a general review of previous work on scene text recognition respectively. While text detection aims to localize text regions in images by filtering out non text outliers from cluttered background text recognition is to transform image-based text information in the detected regions into readable text codes. Scene text recognition is still an open topic to be addressed. In the Robust Reading Competition of International Conference on Document Analysis and Recognition (ICDAR) the best word recognition rate for scene images was only about 41.2%. In general, scene text characters are composed of cross-cutting stroke components in uniform colors and multiple orientations, but they are usually influenced by some font distortions and background outliers.

We observe that text characters from different categories are distinguished by boundary shape and skeleton structure, which plays an important role in designing character recognition algorithm. Current optical character recognition (OCR) systems can achieve almost perfect recognition rate on printed text in scanned documents, but cannot accurately recognize text information directly from camera-captured scene images and videos, and are usually sensitive to font scale changes and background interference which widely exists in scene text. Although some OCR systems have started to support scene character recognition, the recognition performance is still much lower than the recognition for scanned documents. Many algorithms were proposed to improve scene-image-based text character recognition. Weinman et al combined the Gabor-based appearance model, a language model related to simultaneity frequency and letter case, similarity model, and lexicon model to perform scene character recognition. Neumann et al. proposed a real time scene text localization and recognition method based on extremely regions Smith et al. built a similarity model of scene text characters based on SIFT, and maximized posterior probability of similarity constraints by integer programming. It adopted conditional random field to combine bottom-up character recognition and top-down word-level recognition. Here modelled the inner character structure by defining a dictionary of basic shape codes to perform character and word retrieval without OCR on scanned documents. Somebody extracted local features of character patches from an unsupervised learning method associates with a variant of K-means clustering, and pooled them by cascading sub-patch features.

In a complete performance evaluation of scene text character recognition was carried out to design a discriminative feature representation of scene text character structure. In a part-based tree structure model was designed to detect text characters under Latent-SVM and recognize text words from text regions under conditional random field. In Scale Invariant Feature Transform (SIFT) feature matching was adopted to recognize text characters in different languages, and a voting and geometric verification algorithm was presented to filter out false positive matches. In generic object recognition method was imported to extract scene text information. A dictionary of words to be spot is built to improve the accuracy of detection and recognition. Character structure was modelled by HOG features and cross correlation analysis of character similarity for text recognition and detection. In

Random Ferns algorithm was used to perform character detection and constructed a system for query-based word detection in scene images.

In natural scene, most text information is set for instruction or identifier. Text strings in print font are located at signage boards or object packages. They are normally composed of characters in uniform color and aligned arrangement, while non-text Background outliers are in the form of disorganized layouts.Thecolor uniformity and horizontal alignment were employed to localize text regions in scene images. In our current work, scene text detection process is improved to be compatible with mobile applications.

Bag-of-Words Model (Bow)

The BOW model represents a character patch from the training set as a frequency histogram of visual words. The BOW representation is computationally efficient and resistant to intra-class variations.

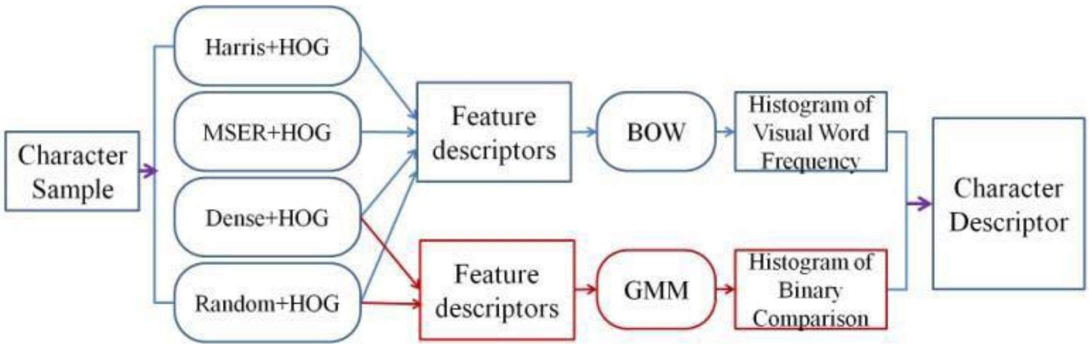

Fig1. *Flowchart of our proposed character descriptor.*

Which combines four key point detectors, and HOG features are extracted at key points. Then BOW and GMM are employed to respectively obtain visual word histogram and binary comparison histogram. At first, k-means clustering is performed on HOG features extracted from training patches to build a vocabulary of visual words. Then feature coding and pooling are performed to map all HOG features from a character patch into a histogram of visual words. We adopt soft-assignment coding and average pooling schemes in the experiments. More other coding/pooling schemes will be tested in our future work. For each of the four feature detectors HD, MD, DD, and RD, we build a vocabulary of 256 visual words. This number of visual words is experimentally chosen to balance the performance of character recognition and the computation cost. At a character patch, the four detectors are applied to extract their respective key points, and then their corresponding HOG features are mapped into the respective vocabularies, obtaining four frequency histograms of visual words. Each histogram has 256 dimensions. Then we cascade the four histograms into BOW-based feature representation in $256 \times 4 = 1024$ dimensions.

IMPLEMENTATION

We have developed demo systems of scene text extraction in Android-Mobile platforms. We integrate the functional modules of scene text detection and text recognition. It is able to detect regions of text strings from cluttered background, and recognize characters in the text regions. Compared with a PC platform, the mobile platform is portable and more convenient to use. Scene text extraction will be more widely used in mobile applications, so it is indispensible to transplant demo system into the popular Android mobile platform. However, two main challenges should be overcome in developing the scene text extraction application in mobile platform .To improve the efficiency, we skip layout analysis of color decomposition in text detection, but directly apply the canny edge map for layout analysis of horizontal alignment. It lowers the accuracy of text detection, but is still reliable for text extraction from nearby object in enough resolutions.

CONCLUSIONS

We have presented a method of scene text recognition from detected text regions, which is compatible with mobile applications. 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 proposed method of scene text recognition outperforms most existing methods. We have also implemented the proposed method to a demo system of scene text extraction on mobile device. The demo system demonstrates the effectiveness of our proposed method in blind-assistant applications, and it also

Proves that the assumptions of color uniformity and aligned arrangement are suitable for the captured text information from natural scene.

REFERENCES

[1] X. Bai, L. J. Latecki, and W.-Y. Liu, “Skeleton pruning by contour partitioning with discrete curve evolution,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 3, pp. 449–462, Mar. 2007.

[2] R. Beaufort and C. Mancas-Thillou, “A weighted finite-state frame work for correcting errors in natural scene OCR,” in Proc. 9th Int. Conf. Document Anal. Recognit., Sep. 2007, pp. 889–893.

[3] X. Chen, J. Yang, J. Zhang, and A. Waibel, “Automatic detection and recognition of signs from natural scenes,” IEEE Trans. Image Process., vol. 13, no. 1, pp. 87–99, Jan. 2004.

[4] A.Coates et al., “Text detection and character recognition in scene images with unsupervised feature learning,” in Proc. ICDAR, Sep. 2011, pp. 440–445.

[5] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2005, pp. 886–893.

[6] T. de Campos, B. Babu, and M. Varma, “Character recognition in natural images,” in Proc. VISAPP, 2009.

[7] B. Epshtein, E. Ofek, and Y. Wexler, “Detecting text in natural scenes with stroke width transform,” in Proc. CVPR, Jun. 2010, pp. 2963–2970.

[8] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, “Object detection with discriminatively trained part-based models,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 9, pp. 1627–1645, Sep. 2010.

[9] T. Jiang, F. Jurie, and C. Schmid, “Learning shape prior models for object matching,” in Proc. CVPR, Jun. 2009, pp. 848–855.

[10] S. Kumar, R. Gupta, N. Khanna, S. Chaudhury, and S. D. Johsi, “Text extraction and document image segmentation using matched wavelets and MRF model,” IEEE Trans. Image Process., vol. 16, no. 8, pp. 2117–2128, Aug. 2007.

[11] L. J. Latecki and R. Lakamper, “Convexity rule for shape decomposition based on discrete contour evolution,” Comput. Vis. Image Understand., vol. 73, no. 3, pp. 441–454, 1999.

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

[13] S. Lu, L. Li, and C. L. Tan, “Document image retrieval through word shape coding,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 30, no. 11, pp. 1913–1918, Nov. 2008.

[14] S. M. Lucas, A. Panaretos, L. Sosa, A. Tang, S. Wong, and R. Young.