Scene Text Detection Based on Enhanced Multi-channels MSER and a Fast Text Grouping Process

Abstract

- Scene text detection has become a popular research in computer vision and pattern recognition field in recent years because of the accurate and rich information carried by scene text.
- Now component-based methods have become the trend, and the detection result is largely determined by the success of filtering text-like non-text regions. The main task of this paper is to reduce the time complexity without a big fall in recall. First an enhanced multi-channels MSER model is introduced:
- Before extracting MSER, the image is sharpened by using the Laplacian and Gaussian blur and multi-channels is utilized, then the step of the threshold used in MSER algorithm is set to the minimum in order to get more refined MSERs. Second, two novel scene text features local contrast and boundary key points are proposed to better distinguish text regions from non-text regions. Finally, a fast text grouping algorithm is achieved which reduce the time complexity. Experiments on both ICDAR 2011 and ICDAR 2013 show that the recall of the proposed method is improved by 3%.

Existing system

- The texts in the natural scene images are with accurate and rich information, and this is very important to image analysis, translation based on images, image searching and so on.
- Over the past two decades, researchers have proposed a number of ways to detect text in natural scene images treat text as connected components which are first extracted by various means, such as color clustering or extreme area extraction, and then the non-text components are filtered using manually designed rules or automatically trained classifiers.

Disadvantages

- Scene text detection has become a popular research in computer vision and pattern recognition field in recent years because of the accurate and rich information carried by scene text.
- Now component-based methods have become the trend, and the detection result is largely determined by the success of filtering text-like non-text regions

Proposed system

- The main task of this paper is to reduce the time complexity without a big fall in recall. First an enhanced multi-channels MSER model is introduced: Before extracting MSER, the image is sharpened by using the Laplacian and Gaussian blur and multi-channels is utilized, then the step of the threshold used in MSER algorithm is set to the minimum in order to get more refined MSERs.
- Second, two novel scene text features local contrast and boundary key points are proposed to better distinguish text regions from non-text regions. Finally, a fast text grouping algorithm is achieved which reduce the time complexity. Experiments on both ICDAR 2011 and ICDAR 2013 show that the recall of the proposed method is improved by 3%

Advantages

- Texture-based methods treat text as a special type of texture and use their texture properties, such as local intensities, filter responses and wavelet coefficients to distinguish between text and non-text regions of an image.
- These methods are usually computationally expensive as all locations and scales should be scanned. In addition, these methods deal mainly with horizontal text and are sensitive to rotation and scale.

Hardware Requirements

- Processor :Intel Pentium IV 1GHz
- RAM :256MB (Min)
- Hard Drive :5GB free space
- Monitor :1024 * 768, High Color inch
- Mouse :Scroll Mouse(Logitech)
- Keyboard :104 keys

Software requirements

- OS : Windows XP/7/8
- Front End : Visual Studio 2010/ netbeans 7.1
- Back End : SQL Server 2005/ heidisql 3.2
- Browser : Any Web Browser

conclusion

- This paper proposes a scene text detection algorithm based on enhanced multi-channels MSER. By sharpening and extracting multiple channels of the image, some text in complex background can be detected, but more non-text regions are introduced as well. In order to filter these nontext regions, this paper introduces two scene text features: local contrast and boundary key points.
- SVM is used to classify the text candidate. Finally, the text grouping algorithm is improved, and the time complexity is reduced to 2 O(nlog n) . Experimentally found, the proposed method has a good performance compared to some similar methods on both ICDAR 2011 and ICDAR 2013 dataset.

References

[1] Wu Y, Shivakumara P, Lu T, et al. Contour Restoration of Text Components for Recognition in Video/Scene Images[J]. IEEE Transactions on Image Processing, 2016, 25: 5622-5634.

[2] Zhu A, Gao R, Uchida S. Could scene context be beneficial for scene text detection? [J]. Pattern Recognition, 2016, 58:204-215.

[3] Yi C, Tian YL. Text string detection from natural scenes by structurebased partition and grouping[J]. IEEE Transactions on Image Processing, 2011, 20(9): 2594–2605

[4] Huang W, Lin Z, Yang J, et al. Text localization in natural images using stroke feature transform and text covariance descriptors[C]//Proceedings of the IEEE International Conference on Computer Vision. 2013: 1241-1248.

[5] Gomez L, Karatzas D. A fast hierarchical method for multi-script and arbitrary oriented scene text extraction[J]. 2014, 19(4):1-15.