2H1-2

Text Localization in Scene Images by Using Character Energy and Maximally Stable Extremal Regions

Jun Baba Akihiro Yamamoto

Graduate School of Informatics, Kyoto University

Text localization in scene images is a challenging problem, and a lot of methods have been proposed for ICDAR dataset. Character Energy is a naïve feature developed by observing images of characters in previous researches. However, the accuracy obtained when Character Energy is used depends on the accuracy of closed boundary detection. In this paper, we propose to combine using Character Energy and detecting Maximally Stable Extremal Regions (MSER). That is, we detect virtually unchanged regions with MSER and compute Character Energy with the boundaries of detected regions. By using our proposal, we would properly extract text locations in scene images of ICDAR dataset.

1. Introduction

Recognizing text in scene images has received a significant amount of attention because it may contribute to improving performance in image search. In contrast to more classical OCR problems [Lin 01], where the characters are typically monotone on fixed backgrounds, text recognition in scene images is potentially far more complicated because it is very difficult to detect exact text locations due to many possible variations in background, lighting, texture and font.

For overcoming this difficulty, a lot of text localization algorithms have been proposed in ICDAR Robust Reading Competition [Lucas 03, Lucas 05, Shahab 11], and there are many researches addressing text localization with using ICDAR dataset [Epshtein 10, Neumann 10, Chen 11]. Most of these existing methods focus on edge features of characters, and Character Energy [Zhang 10] is one of them. Character Energy by looking at naïve features on character boundaries and achived high performance close to the best performance on ICDAR Competition. However, the performance of Character Energy depends on the accuracy of closed boundary detection. If the closed boundary cannot be detected from a character, the method excludes the character from target objects.

In solving this problem, Our idea is to use Maximally Stable Extremal Regions (MSER) [Matas 02]. MSER is an adaptive binarization method to extract low-contrast regions like character regions. In this paper, we propose an efficient boundary object detection which can extract properly character boundaries required in computing Character Energy, by using properties of regions detected by MSER which is referred as 'mser' uncapitalized letters in the rest of this paper. Our algorithm can detect the exact boundaries of a single character as a boundary object even if boundaries extracted by the Canny edge detector are separated and not closed. In this way, our proposed method, combining MSER with Character Energy, can detect more correct text objects composed of detected boundary objects than Zhang's Character Energy method.

The rest of this paper is organized as follows. Section 2 reviews a problem of the previous method and describes our approach to solve the problem. Section 3 describes our boundary object detection in detail. Section 4 demonstrates the performance of our algorithm and section 5 concludes the paper.

Contact: Jun Baba, Graduate School of Informatics, Kyoto University, Yoshida Honmachi, Sakyo-ku, Kyoto, 606-8501, Japan, e-mail: baba@iip.i.kyoto-u.ac.jp

2. Proposed approach

Character Energy represents how likely a boundary object to be a character. It is based on two observations: the boundary of a character is typically closed and the edges of a character typically appear as pairs. With features on character boundaries, we can compute Character Energy based on the observations. In this method, a boundary detected from a character must be closed because, in the case that the detected boundary is not closed, the boundary is not considered as a candidate object. As characters 'N' and '8' in Fig. 1, if a boundary of a character is detected as separate boundaries, the boundary of the each character is considered as other candidate object.

Fig. 1: Examples of character boundaries detected separately

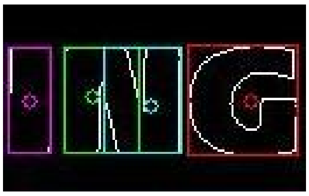
A useful idea for overcoming this problem is that the separated boudaries are likely to compose the same character boundary if their minimal bounding rectangles intersect the same mser. Based on this idea, we regard separated boundary objects as a candidate boundary object by merging the separated objects that their rectangles intersect the same mser. In the following section, we show our object detection algorithm in detail.

3. Character boundary object extraction with MSER detection

First, we detect edges from an input image by using the Canny edge detector. All detected edges are regarded as candidate boundary objects whether the edges are closed or not. In Fig. 2(a), examples of detected boundary objects are shown. MSERs are detected from the input image in parallel with edge detection. As preprocessing of MSER detection, we smooth the input image excluding edge points with a moving average filter in order to make the contrast of character regions lower. Then we sharpen the smoothed image by unsharp masking to raise the contrast of edge points. This process helps us to detect msers efficiently.

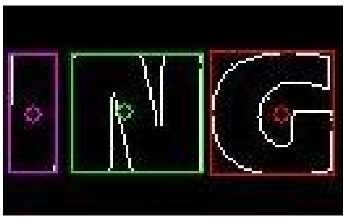
Secondly, we investigate whether each boundary object extracted by the edge detector intersects each mser detected from the preprocessed image (Fig. 2(a) and (b)). More precisely, we look whether the minimum bounding rectangle of a boundary object intersects the minimum bounding rectangle of a mser. Then, to the boundary objects, we find the most similar mser from intersected msers on the viewpoint of properties such as size and aspect ratio. Thus, The most similar mser is chosen as a mser which the boundary object has in its region.

We can see that boundary objects which have the same mser in their regions are likely to compose a closed boundary of a candidate character region. By merging the objects which have the same mser, a closed boundary of a candidate region can be detected properly (Fig. 2(c)).



(a) Rectangles and pixels of boundaries detected separately

(b) Rectangles of 'msers': regions detected by MSER.



(c) Rectangles and pixels of merged boundaries

Fig. 2: Boundary object detection with MSER

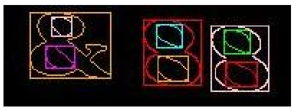
Finally, we construct a inclusion relationship to detect the exact boundary of a single character that contains holes, like characters '8' and '&' in Fig. 3. A inclusion relationship means a relationship between an object that includes other objects fully and the included objects. For example, an external pink boundary in the right '8' in Fig. 3(a) is a larger object and internal green boundary and red boundary in the same '8' are smaller objects that included the larger object fully. As seen in Fig. 3(a), we can see that larger objects and smaller which have similar size and aspect ratio compose a same character boundary. Hence, we merge such similar objects in inclusion relationships if

$$\frac{\min(S(l), S(s))}{\max(S(l), S(s))} \leq t_S \quad \text{and} \quad \frac{\min(A(l), A(s))}{\max(A(l), A(s))} \leq t_A,$$

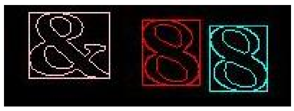
where l and s mean a larger object and a smaller object, and S and A are the size and the aspect ratio of parent object respectively. Thresholds t_S and t_A present with weights w_i ($i = 1, \ldots, 4$) as

$$t_S = \exp(-w_1 * \max(S(l), S(s))/\min(S(l), S(s)) + w_2$$
$$t_A = \exp(-w_3 * \max(A(l), A(s))/\min(A(l), A(s)) + w_4.$$

In this way, we can detect the exact boundary object of a character candidate region (Fig. 3(b)). In addition to our algorithm, we compute the character energy and link energy of the detected boundary objects for detecting text objects. We can eventually obtain text objects by thresholding each energy of the boundary objects.



(a) Rectangles and pixels of Larger objects and smaller



(b) Rectangles and pixels of merged boundary objects

Fig. 3: Merging larger objects and smaller with inclusion relationships

4. Experimental Results

In order to validate the performance of our proposed algorithm, we applied it ICDAR 2003 text locating dataset and use the metrics defined in [[Lucas 03]]. ICDAR 2003 test dataset has 249 indoor and outdoor images. The precision and recall metrics are defined as $p = \sum_{r_e \in E} m(r_e, T)/|E|$ and $r = \sum_{r_t \in T} m(r_t, E)/|T|$, where $m(r, R)$ is the best match for a rectangle r in a set of rectangles R, and E and T are our estimated rectangles and the ground truth rectangles respectively.

Figure 4 shows some detection results of our method. Character objects and text objects are bounded by red boxes and green boxes respectively. We can see that each single character object is correctly detected by our algorithm. We show the performance of our algorithm and previous methods using ICDAR dataset in Table 1. Our algorithm achive higher recall than the previous methods, but the precision of our algorithm is lower than them. We conjecture that it is caused by merging even non-text objects in merging process with msers or conclusion relationships

Fig. 4: Detection results of our method

Algorithm	precision	recall
Our algorithm	0.60	**0.75**
Zhang [Zhang 10]	0.62	0.73
Chen [Chen 11]	0.73	0.60
Epshtein [Epshtein 10]	0.73	0.60
Kim [Shahab 11]	0.83	0.62
Neumann [Neumann 10]	0.64	0.72

Table 1: Evaluation of text detection algotirhms

The performance of our method still depends on the accuracy of the Canny edge detector as well as the previous Character Energy method. However, this dependency can be decreased by integrating edge images detected from various grayscale images.

5. Conclusion

In this paper, we propose an efficient boundary object detection with properties of msers, for extracting properly character boundaries required in computing Character Energy. Our algorithm merges boundaries detected separately into a single boundary, by analyzing the rectangle overlaps of boundaries and msers and using inclusion relationships. We evaluated our boundary detection algorithm on ICDAR text locating dataset 2003. The result demonstrates that our method achives the higher recall performance than the existing algorithms in [Chen 11, Epshtein 10, Neumann 10]. As a future work, we will integrate edge images detected from various grayscale images in order to improve the edge detection.

References

[Chen 11] H. Chen, S. S. Tsai, G. Schroth, D. M. Chen, R. Grzeszczuk and B. Girod, "Robust text detection in natural images with edge-enhanced Maximally Stable Extremal Regions," in *image Processing (ICIP), 2011 18th IEEE International Conference on*, 11-14 Sept. 2011, pp.2609-2612.

[Epshtein 10] B. Epshtein, E. Ofek, and Y. Wexler, "Detecting text in natural scenes with stroke width transform," in *CVPR 2010*, pp. 2963-2970.

[Lin 01] X. Lin, "Reliable OCR solution for digital content remastering," in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, 2001.

[Lucas 03] S. M. Lucas, A. Panaretos, L. Sosa, A. Tang, S. Wong and R. Young, "ICDAR 2003 Robust Reading Competitions," in *Proceedings of the Seventh International Conference on Document Analysis and Recognition*, 2003, pp.682-687.

[Lucas 05] S. M. Lucas, "ICDAR 2005 text locating competition results," in *ICDAR*, 2005, vol. I, pp. 80-85.

[Matas 02] M. S. Extremal, J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust Wide Baseline Stereo from Maximally Stable Extremal Regions," in *British Machine Vision Conference*, 2002, pp. 384-393.

[Neumann 10] L. Neumann and J. Matas, "A method for text localization and recognition in real-world images," in *ACCV 2010*,

[Shahab 11] A. Shahab, F. Shafait, and A. Dengel, "ICDAR 2011 Robust Reading Competition Challenge 2: Reading Text in Scene Images," in *Int. Conf. on Document Analysis and Recognition (ICDAR) 2011*, Sep. 2011, pp. 1491-1496. ser. LNCS 6495, vol. IV, Nov. 2010, pp. 2067-2078.

[Zhang 10] J. Zhang and R. Kasturi, "Character energy and link energybased text extraction in scene images," in *ACCV 2010*, ser. LNCS 6495, vol. II, Nov. 2010, pp. 832-844.