Symmetry-Based Text Line Detection in Natural Scenes

Zheng Zhang[1], Wei Shen[2], Cong Yao[1], Xiang Bai[1]
[1] School of EIC, Huazhong University of Science and Technology. [2] Key Lab of Specialty Fiber Optics and Optical Access Networks, Shanghai University.

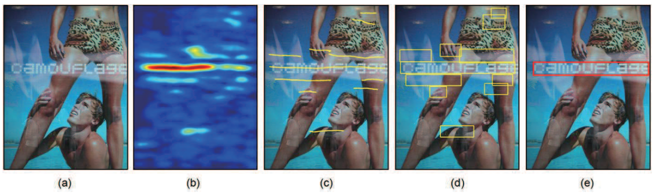

Figure 1: Schematic pipeline of our symmetry-based text-line detection algorithm. (a) Input image; (b) Response map of the symmetry detector; (c) Symmetrical point grouping; (d) Estimated bounding boxes based on the detected symmetrical axes. (e) Detection result after false alarm removal.

There are mainly two classes of mainstream methods for scene text detection: those based on a sliding window [1, 5] and those based on connected component extraction [2, 7, 8]. The latter category has become the mainstream in the field of scene text detection, since these methods are usually more efficient and relatively insensitive to variations in scale, orientation, font, and language type. In these methods, Maximally Stable Extremal Regions (MSER) [4] and Stroke Width Transform (SWT) [2] are widely adopted as the basic representation due to their efficiency and stability. However, such representations may perform poorly under severe conditions, such as blur, non-uniform illumination, low resolution and disconnected strokes.

Different from individual characters, text lines always bear distinctive symmetry and self-similarity properties. The symmetry property of text lines comes from both themselves and their local backgrounds. Taking advantage of this property, we approach the text detection problem from another perspective and propose to seek the symmetrical axes of text lines in natural images, with a symmetry detector.

The pipeline of the proposed symmetry-based text line detection approach is shown in Fig. 1. For each pixel in the image the probability of being on the symmetrical axis of a text line is estimated using a predefined symmetry template (see Fig. 2) at first. Then, text line candidates are formed by grouping the pixels on symmetrical axes and estimating their corresponding bounding boxes. Finally, false positives (non-text candidates) are identified and eliminated with CNN classifiers [3]. To deal with texts of different sizes, the above described procedure is performed at multiple scales. Detection activations from different scales are merged and non-maximum suppression is adopted to remove redundant detections.

Symmetry-Based Text Line proposals

Symmetry Template. We devise a symmetry template that is suitable for seeking symmetrical structures, following [6]. The template, as illustrated in Fig. 2, consists of four rectangles with equal size $s \times 4s$, denoted by R_T, R_{MT}, R_{MB} and R_B, respectively. The rectangle formed by the two middle rectangles R_{MT} and R_{MB} are denoted by R_M. The height of each rectangle, i.e s, is defined as the scale of the template.

Features. To detect symmetry axes as text line proposals, we employ two types of features: symmetry feature and appearance feature, which capture the intrinsic properties of text. Assume that the template is centered at location (x,y) on the image plane and let $h^c_{x,y}(R_P) (P \in \{T,M,B,MT,MB\})$ denote the histogram of the low-level image cue c in the rectangle R_P.

Symmetry feature is used to characterize the self-similarity and symmetry property of character groups. Character groups have self-similarity since adjacent characters bear similar color and structure. The self-similarity is defined as the difference between the two middle rectangles in low-level image cues:

$$S^c_{x,y} = \chi^2(h^c_{x,y}(R_{MT}), h^c_{x,y}(R_{MB})), \tag{1}$$

where $\chi^2(\cdot)$ is the χ^2-distance function.

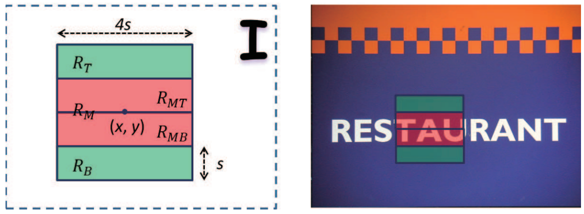

Figure 2: Template used to compute the features for symmetry detection.

Meanwhile, a text region is usually highly dissimilar to its local background. This can be seen as another kind of symmetry, since the contents in the middle rectangles (R_{MT} and R_{MB}) are both different from those in the outer rectangles (R_T and R_B). To measure this dissimilarity, we define the contrast feature as the differences of the low-level image cues within the rectangle pairs:

$$Ct^c_{x,y} = \chi^2(h^c_{x,y}(R_T), h^c_{x,y}(R_{MT})), \tag{2}$$

and

$$Cb^c_{x,y} = \chi^2(h^c_{x,y}(R_B), h^c_{x,y}(R_{MB})). \tag{3}$$

The symmetry feature is effective at finding text lines in images, but it also fires on some non-text symmetrical structures. To better distinguish text and non-text, we employ appearance feature, as it has been widely used in previous works [1]. Specifically, the local binary pattern (LBP) feature of the middle rectangle R_M is taken as the appearance feature.

Symmetry Axis Detection. For symmetry axis detection, we train a strong classifier to estimate the probability of being on a symmetry axis at each pixel. Random Forest is chosen as the classifier for its high efficiency and performance. To train the symmetry axis detector, the ground truth rectangles of text lines are required. However, the current text detection benchmarks, such as ICDAR 2011 and 2013, only provide bounding boxes that correspond to parts of text. To produce text line level ground truth, we simply compute the center lines of the bounding boxes.

False Positive Removal

A portion of the text candidates generated in the proposal generation stage are non-text. The purpose of false positive removal is to identify and eliminate such non-text candidates. Inspired by the deep learning methods of Jaderberg *et al.* [3], we also adopt CNN classifiers for false positive removal. Different from [3], which only used CNN classifier for patch or character level discrimination, we train two classifiers that work at character level and text region level, respectively.

[1] X. Chen and A. Yuille. Detecting and reading text in natural scenes. In *Proc. of CVPR*, 2004.

[2] B. Epshtein, E. Ofek, and Y. Wexler. Detecting text in natural scenes with stroke width transform. In *Proc. of CVPR*, 2010.

[3] M. Jaderberg, A. Vedaldi, and A. Zisserman. Deep features for text spotting. In *Proc. of ECCV*, 2014.

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

[5] L. Neumann and J. Matas. Scene text localization and recognition with oriented stroke detection. In *Proc. of ICCV*, 2013.

[6] S. Tsogkas and I. Kokkinos. Learning-based symmetry detection in natural images. In *Proc. of ECCV*, 2012.

[7] C. Yao, X. Bai, W. Liu, Y. Ma, and Z. Tu. Detecting texts of arbitrary orientations in natural images. In *Proc. of CVPR*, 2012.

[8] C. Yao, X. Bai, and W. Liu. A unified framework for multi-oriented text detection and recognition. *IEEE Trans. Image Processing*, 23(11): 4737–4749, 2014.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.