

Scene Character Detection by an Edge-Ray Filter

Rong Huang*, Palaiahnakote Shivakumara[†] and Seiichi Uchida*

**Kyushu University, Fukuoka, Japan*

[†]*University of Malaya, Kuala Lumpur, Malaysia*

Abstract—Edge is a type of valuable clues for scene character detection task. Generally, the existing edge-based methods rely on the assumption of straight text line to prune away the non-character candidates. This paper proposes a new edge-based method, called edge-ray filter, to detect the scene character. The main contribution of the proposed method lies in filtering out complex backgrounds by fully utilizing the essential spatial layout of edges instead of the assumption of straight text line. Edges are extracted by a combination of Canny and Edge Preserving Smoothing Filter (EPSF). To effectively boost the filtering strength of the designed edge-ray filter, we employ a new Edge Quasi-Connectivity Analysis (EQCA) to unify complex edges as well as contour of broken character. Label Histogram Analysis (LHA) then filters out non-character edges and redundant rays through setting proper thresholds. Finally, two frequently-used heuristic rules, namely aspect ratio and occupation, are exploited to wipe off distinct false alarms. In addition to have the ability to handle special scenarios, the proposed method can accommodate dark-on-bright and bright-on-dark characters simultaneously, and provides accurate character segmentation masks. We perform experiments on the benchmark ICDAR 2011 Robust Reading Competition dataset as well as scene images with special scenarios. The experimental results demonstrate the validity of our proposal.

I. INTRODUCTION

As a type of descriptive feature, characters appearing in the natural scene images usually provide crucial semantic information. However, unlike the document-based image analysis, detecting scene characters is full of challenges due to the unconstrained character appearances, namely variations in font, size, color, layout and location.

In the last decade, a great number of impressive techniques [1–12] have been elaborately designed to narrow the gap between the limited capability of Optical Character Recognition (OCR) and the unconstrained appearances of scene characters. Among the existing techniques, edge-based methods have attracted extensive attention, since they utilize the essential and reliable contrast feature, and enable working free from an exhaustive scan of the original image. Chen *et al.* [2] probed the characters based on the fact that strokes contain cross edges. Each normalized text line was segmented into multiple binary images for the subsequent confidence calculation. To extract the stroke width, Subramanian *et al.* [3] sought a set of dashes by analyzing a group of horizontal scanning lines. The method proposed by Dinh *et al.* [4] traversed all horizontal lines to find stable valleys between double edge pixels. Then, a stroke width-based adaptive morphological dilation operator was exploited to connect edge pixels into candidate regions. Zhang *et al.* [5] calculated character energy based on the similarity of stroke edges, and clustered character regions into text line through the link energy. Based on the inverse response between Sobel and Canny, Shivakumara *et al.* [6] employed separate thresholds for high and low images to detect horizontal text lines. Epshtein *et al.* [9] invented

a Stroke Width Transform (SWT) operator which mapped each grayscale pixel to a stroke width value. Then, Connected Component (CC) analysis was applied to gather the pixels with similar stroke width value. Heuristic rules as well as the assumption of horizontal text line were employed to remove non-character CCs. Recently, Yao *et al.* [10] extended SWT by combining the component level and chain level classifier to accommodate the text appearing with arbitrary orientations.

However, most previous methods filtered out non-character elements based on the assumption that scene characters appear along a straight text line. Practically, in natural scene images, there exist numerous special scenarios such as single character, characters along with curve or more complex layouts as shown in Fig.1 (b)~(e). Moreover, for these existing methods, direct extension to handle all above scenarios is not a trivial matter.

In this paper, inspired by the ray shooting stage of SWT [9], we develop a novel character-level edge-ray filter characterized by the strong filtering strength for suppressing the redundant detections, so that relatively pure detections can be achieved even without the aid of the assumption of straight text line. The proposed edge-ray filter consists of four modules: (A) edge detector, (B) Quasi-Connectivity Analysis (EQCA) module, (C) Label Histogram Analysis (LHA) module and (D) verification module. To relieve from the toil of threshold tuning, we design a tractable edge detector, which combines Canny operator with Edge Preserving Smoothing Filter (EPSF) [13]. The EQCA module assigns a unique label to each quasi-connected edge, and eliminates tiny noise components. After that, each surviving edge pixel shoots bidirectional rays along its gradient direction and the corresponding opposite direction. For each quasi-connected edge, LHA module creates label histograms for endpoints, and then removes the edge or its rays if the label distribution does not satisfy regulations. In the final verification module, we wipe off distinct false alarms by using aspect ratio and occupation. It is worthwhile to point out that the output image is an accurate character segmentation mask instead of loose bounding box covers, which allows us to directly feed the output image into a subsequent OCR module.

We evaluate the performance of the proposed edge filter on the benchmark ICDAR 2011 Robust Reading Competition dataset [14]. Experimental results demonstrate that our method can cope with the special scenarios, and yields a remarkable performance in spite of abandoning the text line assumption.

II. METHODOLOGY

A. Edge Detector

Numerous previous works [4, 9, 10] employed the Canny operator to detect the distinct boundaries between foreground and background. However, threshold setting for Canny operator is not a straightforward work. For preserving character edge pixels as much as possible, the double thresholds of Canny operator should be low enough, while the low thresholds cause

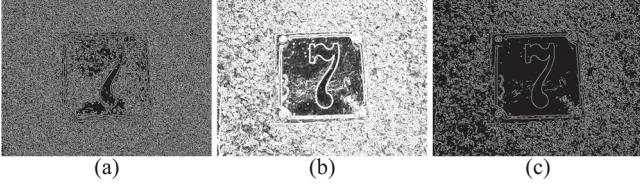


Fig. 2: Resulting images of the edge detector. (a) $C(x, y)$. (b) $G(x, y)$. (c) $E(x, y)$.

a mass of noise-like detections as shown in Fig.2 (a) (see the original image in Fig.1 (b)). To alleviate this dilemma, we combine the Canny operator with Edge Preserving Smoothing Filter (EPSF) [13] towards a tractable edge detector.

Edge preserving smoothing filter, as the name implies, allows us to reduce noise without affecting edge pixels. The resulting magnitude map $G(x, y)$ is normalized to the range of grayscale as shown in Fig.2 (b). See [13] for details. Then, the combined edge image $E(x, y)$ takes the following form:

$$E(x, y) = \begin{cases} 1 & \text{if } C(x, y) = 1 \text{ and } G(x, y) > T_G \\ 0 & \text{otherwise} \end{cases}$$

where $C(x, y)$ stands for the resulting image of Canny operator. The threshold T_G for $G(x, y)$ is automatically determined by Otsu's method. As displayed in Fig.2 (c), the combined detector can suppress a great deal of texture edge pixels.

B. Edge Quasi-Connectivity Analysis Module

To effectively boost the filtering strength of edge-ray filter, we propose an Edge Quasi-Connectivity Analysis (EQCA) which has the ability to connect edges that are not directly adjacent to each other. For some high-frequency scenes like lawn, foliage or brick wall, EQCA is capable of unifying their cluttered edges. This process renders the subsequent LHA a higher probability to wholly reject these non-characters but high-frequency contents. Moreover, as a by-product, EQCA reconstructs the broken character contours by tagging them with the same label.

EQCA starts with the general connectivity analysis. Let A denote an adjacency relation of edge pixels in $E(x, y)$. We consider 8-neighbour connectivity in this paper so that all eight peripheral pixels of a 3×3 patch are adjacent to the central pixel. Thus, a completed edge connected subset P is formed by the following two conditions.

- (1) $\forall p_i, p_j \in P, \exists$ a contiguous pathway $p_i A p_{i+1} A \dots A p_j$.
- (2) $\forall p_i \in P$ and $\forall p_j \notin P, \nexists$ a contiguous pathway.

Each subset P_m is then labeled an index number m . We eliminate the subset if $L_m < 10$, where L_m stands for the total number of edge pixels tagged with m . Figure 3 (a) shows a resulting labeled edge map $E'(x, y)$, in which edge pixels are coated with different colors according to their labels.

In the following, we describe the procedures of EQCA.

- (1) Scan the $E'(x, y)$, and collect label pairs at each edge gap. The label pair takes the form $\{m, n\}$ normalized so that $m < n$ holds. See some examples in Fig.4.
- (2) Put the label pair which occurs more than T_N times into a list SC .
- (3) For each label pair $\{m, n\} \in SC$, iteratively generate a label union $U(\{m, n\})$ as follows. Initially, $U(\{m, n\}) = \{m, n\}$. Traverse the list SC , if $U(\{m, n\}) \cap \{m_i, n_i\} \neq \emptyset$, where $\{m_i, n_i\} \in SC$, update the label union through $U(\{m, n\}) = U(\{m, n\}) \cup \{m_i, n_i\}$.
- (4) Unify the subsets P_{m_k} , where $m_k \in U(\{m, n\})$, by assigning a new uniform label.

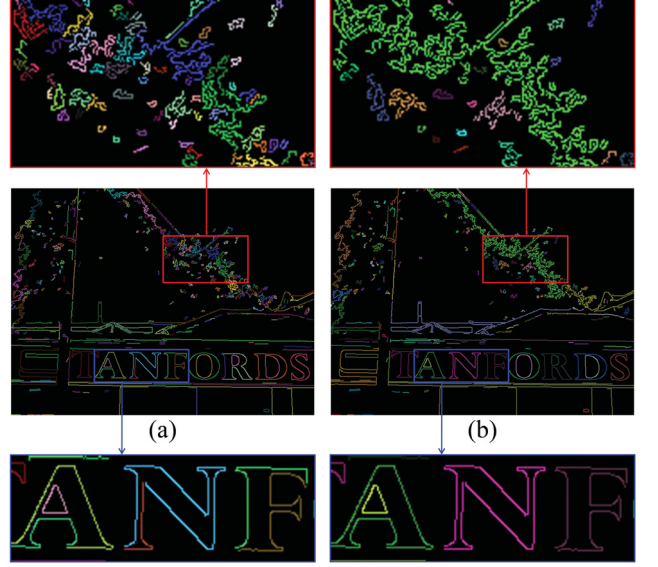


Fig. 3: (a) $E'(x, y)$. (b) $E''(x, y)$.

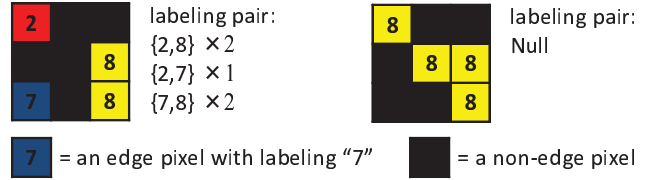


Fig. 4: Examples of labeling pair.

The proposed EQCA connects the physically separate subsets if they share more than T_N identical label pairs. Apparently, the strength of edge fusion will decrease when T_N gets larger. Under the extreme case $T_N \rightarrow +\infty$, the EQCA degenerates into the general connectivity analysis. In this paper, we fix $T_N = 2$ empirically.

A new labeled edge image $E''(x, y)$ is given in Fig.3 (b). In the regions cut by red and blue squares, the edges of foliage as well as broken characters are coated with the same color at this stage, indicating that they have the same label.

C. Label Histogram Analysis Module

In this section, we elaborate the details of the proposed Label Histogram Analysis (LHA) which is necessary for removing redundant edges and rays. The explicit task of LHA is to filter out non-character edges as well as the redundant rays, and the working principle is based on the observation that spatial layout of edges between inner and outer of a character reveals disparate characteristics. Label histogram can exactly represent this characteristics.

For completeness, we start from the ray shooting stage which imitates the SWT [9]. As illustrated in Fig.5, each edge pixel p_i in P_m shoots bidirectional rays along its gradient direction d_{p_i} and the corresponding opposite direction $-d_{p_i}$. A pair of bidirectional rays has the following form:

$$rA_i = p_i + nA \cdot d_{p_i} \quad \text{and} \quad rB_i = p_i - nB \cdot d_{p_i}$$

Then, two rays are lengthened by increasing $nA, nB > 0$ until they reach the endpoints qA_i and qB_i , respectively. An endpoint is preserved if its gradient direction is roughly opposite to the ray direction, concretely

$$d_{qA_i} = -d_{p_i} \pm \pi/2, \quad d_{qB_i} = d_{p_i} \pm \pi/2 \quad (1)$$

Otherwise the endpoint will be labeled "null". Note that if a ray does not find an endpoint no matter how long it lengthens,

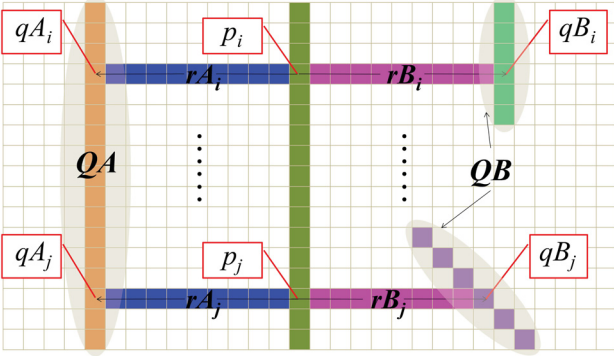


Fig. 5: Diagram of shooting rays.

qA_i or qB_i will be tagged with the label “null” as well.

To better describe the LHA in a concise matter, we give some definitions of symbols as well as terminologies.

$QA(QB)$: a subset collects the endpoints $qA_i(qB_i)$, where $i = 1, \dots, L_m$. See the diagram in Fig.5.

LHQA(LHQB): short for the label histogram of $QA(QB)$.

$LQA(LQB)$: the label set of $QA(QB)$ recording the labels of endpoints in the subset $QA(QB)$.

l_m : the number of the endpoints with label m in the subset $QA(QB)$.

Stroke ray: the ray which covers the character stroke.

Redundant ray: non-stroke ray.

For each subset P_m , its two histograms LHQA and LHQB are created based on the label distribution of QA and QB , respectively. Figure 6 takes the region around character “O” as example. In order to simplify the exposition, we manually assign the artificial labels, namely index numbers $1 \sim 7$, to every subsets. For the outer ring of character “O”, namely the subset P_1 , its two label sets will be $LQA = \{\text{null}, 7\}$ and $LQB = \{\text{null}, 2, 3, 4, 5, 6\}$. Likewise, for the inner ring P_7 , the label sets are $LQA = \{\text{null}, 1\}$ and $LAB = \{\text{null}, 7\}$.

The comparison between P_1 's LHQA and LHQB reveals that the number of stroke rays usually occupies a dominating proportion, while the label distribution of redundant rays takes on a scattered fashion. On the basis of this observation, we analyze the label histograms through following three regulations: *Regulation I* Maximum ratio regulation, *Regulation II* Minimum ratio regulation and *Regulation III* Average ratio regulation. Moreover, *Regulation IV* Average distance regulation is specially designed to handle the histograms distributing like P_7 's LHQA and LHQB.

The first maximum ratio regulation can be expressed by four equations below.

$$\begin{aligned} Max_{AM} &= \frac{\max_{k \in LQA} l_k}{L_m}, & Max_{BM} &= \frac{\max_{k \in LQB} l_k}{L_m} \\ Max_{AK} &= \frac{\max_{k \in LQA} l_k}{L_{\text{argmax}_{k \in LQA} l_k}}, & Max_{BK} &= \frac{\max_{k \in LQB} l_k}{L_{\text{argmax}_{k \in LQB} l_k}} \end{aligned}$$

Go back to the example of P_1 in Fig.6, we obtain the corresponding four results as follows: $Max_{AM} = l_7/L_1$, $Max_{BM} = l_4/L_1$, $Max_{AK} = l_7/L_7$ and $Max_{BK} = l_4/L_4$.

A binary decision is made by setting a threshold T_{MAX} . Specifically, if $Max_{AM}(Max_{BM}) < T_{MAX}$ or $Max_{AK}(Max_{BK}) < T_{MAX}$, the rays between P_m and $QA(QB)$ are discarded. Moreover, the edge subset P_m itself will be wiped off if all its bidirectional rays are discarded. Compare the difference between Fig.7 (b) and (c).

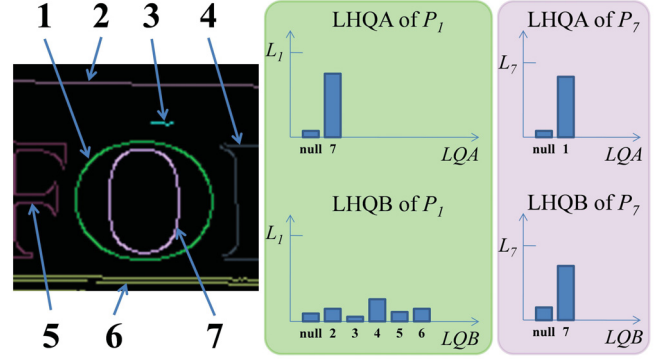


Fig. 6: An example of the label histograms.

To suppress the abrupt redundant rays, the second minimum ratio regulation is designed as follows.

$$\begin{aligned} Min_{AM} &= \frac{\min_{k \in \{LQA-\text{null}\}} l_k}{L_m}, & Min_{BM} &= \frac{\min_{k \in \{LQB-\text{null}\}} l_k}{L_m} \\ Min_{AK} &= \frac{\min_{k \in \{LQA-\text{null}\}} l_k}{L_{\text{argmin}_{k \in \{LQA-\text{null}\}} l_k}}, & Min_{BK} &= \frac{\min_{k \in \{LQB-\text{null}\}} l_k}{L_{\text{argmin}_{k \in \{LQB-\text{null}\}} l_k}} \end{aligned}$$

Likewise, for the P_1 in Fig.6, we have $Min_{AM} = l_7/L_1$, $Min_{BM} = l_7/L_7$, $Min_{AK} = l_3/L_1$ and $Min_{BK} = l_3/L_3$. We filter out the corresponding rays with minimum ratio, if $Min_{AM}(Min_{BM}) < T_{MIN}$ or $Min_{AK}(Min_{BK}) < T_{MIN}$, where T_{MIN} is a minimum threshold. See the effect of *Regulation II* by comparing Fig.7 (c) and (d). Note that the minimum ratio regulation does not affect the edge subset P_m and only targets on the abrupt rays, otherwise some innocent character edges may be improperly discarded.

To further increase the reliability, we regulate the following average ratio measurements.

$$\begin{aligned} Avg_{AM} &= \frac{\sum_{k \in LQA} l_k/L_m}{|LQA|}, & Avg_{BM} &= \frac{\sum_{k \in LQB} l_k/L_m}{|LQB|} \\ Avg_{AK} &= \frac{\sum_{k \in LQA} l_k/L_k}{|LQA|}, & Avg_{BK} &= \frac{\sum_{k \in LQB} l_k/L_k}{|LQB|} \end{aligned}$$

where the operator $|\cdot|$ counts the number of components in a set. The threshold can be set as the average of T_{MAX} and T_{MIN} . Similarly, if $Avg_{AM}(Avg_{BM})$ or $Avg_{AK}(Avg_{BK})$ is less than the average threshold, the rays between P_m and $QA(QB)$ are discarded. The subset P_m itself will be abandoned if the *Regulation III* does not hold for both QA and QB . The average ratio regulation has the ability to remove the cluttered edge whose label histogram generally has a prominent peak accompanied by too many short columns. The power of *Regulation III* is exhibited in Fig.7 (d) and (e).

The final average distance regulation enables us to eliminate the annoying rays which fill the hole of character. Towards this end, we first find the midpoint of each ray, then extract the bounding box BB_m of P_m as well as its attributes, namely height h_m , width w_m and centroid point C_m . The average distance is defined as follows.

$$Avg_{disA} = \frac{\sum_{i=1}^{|QA|} \|\vec{C}_m - c_{rA_i}\|}{|QA|}, \quad Avg_{disB} = \frac{\sum_{i=1}^{|QB|} \|\vec{C}_m - c_{rB_i}\|}{|QB|}$$

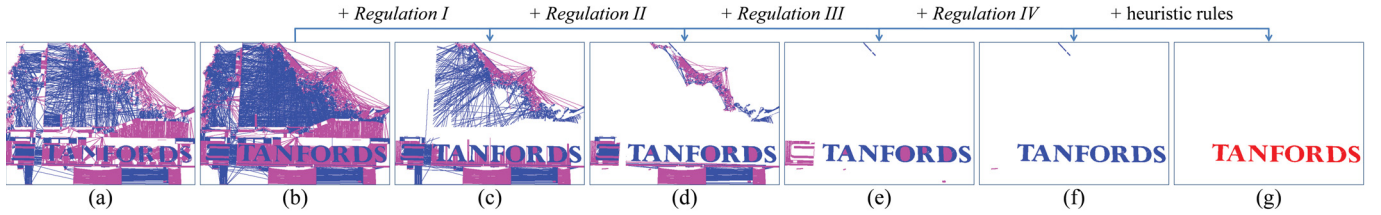


Fig. 7: The evolution of resulting image.

where $c_{rA_i}(c_{rB_i})$ denotes the midpoint of ray $rA_i(rB_i)$, while $\|\cdot\|$ computes the magnitude of a vector. Take the inner ring of “O” (P_7 in Fig.6) as example. The estimated average distance of the stroke rays Avg_{disA} is greater than $\text{MAX}(h_7, w_7)/2$, while for the redundant rays, $Avg_{disB} \approx 0$. Although here two values show a distinct relation $Avg_{disA} > Avg_{disB}$, it is not true for all the character edges. The subset P_1 gives an immediate counterexample. Therefore, making decision by directly comparing Avg_{disA} and Avg_{disB} may ruin innocent character edges.

To solve this dilemma, we design the decision criterion without introducing new thresholds as follows.

$$Avg_{disA} < (Max_{BM} + Max_{BK}) \cdot \max(h_m, w_m)/4 \quad (2)$$

$$Avg_{disB} < (Max_{AM} + Max_{AK}) \cdot \max(h_m, w_m)/4 \quad (3)$$

If equation (2)(or (3)) holds, the rays between P_m and $QA(QB)$ will be removed. We analyze the mechanism through the same example. For the inner ring P_7 , according to the *Regulation I*, $(Max_{BM} + Max_{BK})/2$ and $(Max_{AM} + Max_{AK})/2$ are both close to 1. On the other hand, as just mentioned, $Avg_{disA} > \max(h_7, w_7)/2$ and $Avg_{disB} \approx 0$ so that only the redundant rays between P_7 and its QB will be removed. On the contrary, for the outer ring P_1 , although Avg_{disA} of stroke rays is slightly less than $\max(h_1, w_1)/2$, its $(Max_{BM} + Max_{BK})/2$ is much less than 1. Therefore, the edge subset P_1 as well as its stroke rays are preserved.

It is worthwhile to highlight that for all four regulations, only two parameters T_{MAX} and T_{MIN} are required to tune, which represents the simplicity of parameter setting. In this paper, we set $T_{MAX} = 0.45$ and $T_{MIN} = 0.05$ experimentally.

D. Verification Module

To further filter out some distinct but stubborn false alarms, two popular heuristic rules, aspect ratio and occupation ratio, are applied under relaxed parameter setting. Specifically, a CC will be removed if (i) its aspect ratio is greater than 10 or less than 0.1. Or (ii) its occupation ratio is beyond the range [0.1,0.9]. The final result is given in the form of the red character segmentation mask as shown in Fig.7 (g).

The LHA fully utilizes the spatial layout of edges so that the proposed edge-ray filter can remarkably remove most redundant rays even without the aid of the assumption of straight text line. Since the direction constraint in equation (1) is relaxed from $\pm\pi/6$ to $\pm\pi/2$ compared with the original setting in [9], our proposal returns a more accurate segmentation mask in the sense that the rays densely cover the curved and joint parts of a stroke. The contrast between Fig.7 (a) and (b) indicates that although the original conservative setting can naturally exclude a considerable number of redundant rays, the sparse cover on characters leads to the incomplete segmentation mask which may severely attenuate the subsequent recognition stage. In addition, benefiting from shooting bidirectional rays, the proposed method can detect both dark-on-bright and bright-on-dark characters simultaneously. The above merits are further demonstrated by the following experiments.

TABLE I: Quantitative detection results. Note that only our method is free from the assumption of the straight text line.

| Method | r (%) | p (%) | F (%) |
|---------------------------------------|--------------|--------------|--------------|
| Kim’s Method | 62.47 | 82.98 | 71.28 |
| Yi’s Method | 58.09 | 67.22 | 62.32 |
| TH-TextLoc System | 57.68 | 66.97 | 61.98 |
| Edge-Ray Filter (with EQCA) | 63.77 | 61.98 | 62.86 |
| Edge-Ray Filter (without EQCA) | 60.24 | 50.03 | 54.66 |

III. EXPERIMENT

In this section, we exhibited the validity of the proposed edge-ray filter. The performance was evaluated on the ICDAR 2011 Robust Reading Competition test dataset [14] with respect to three canonical measures, namely recall r , precision p and F-measure F . We conducted two comparative studies. The first one was between three top-ranked methods listed in [14]. The second one was conducted with the intent to determine the importance of EQCA.

The detailed quantitative results listed in Table I demonstrated that the edge-ray filter could achieve the remarkable performance which was comparable with the top-ranked methods. Particularly, our r ranked the first in the competition results, indicated that the proposed method had great ability to seek the scene characters. On the other hand, the p was still acceptable even without the aid of the assumption of straight text line.

Then, the second comparative study had the EQCA turned off. All three measures deteriorated dramatically. The precision dropped about 10%, which suggested that EQCA indeed increased the filtering strength.

Five examples were given in Fig.1. The four natural scene images with special scenarios (b)~(e) were harvested from the Internet, and we manually prepared their ground-truth images. To clearly exhibit the detection results, each correctly detected character was colored red, while the missing ones were represented by black. In addition, green pixels stood for the redundant detections. As expected, the edge-ray filter could detect dark-on-bright and bright-on-dark characters simultaneously (see (a)), and accurately coped with the special scenarios like single character (see (b)), curve text line (see (c)), and complex layouts (see (d) and (e)).

To reflect the filtering strength of the edge-ray filter, Table II listed the evolution of the total number of (quasi-)connected edges. Two measurements AFR_{QECA} and AFR_{LHA} computed the filtering ratio averaged over all test images. Note that, as described before, QECA just fused the cluttered edges instead of directly pruning them.

Since the proposed method is greatly dependent on the edge map, it is an undeniable fact that the edge-ray filter still shares a common problem of the edge-based methods. Obstacles with

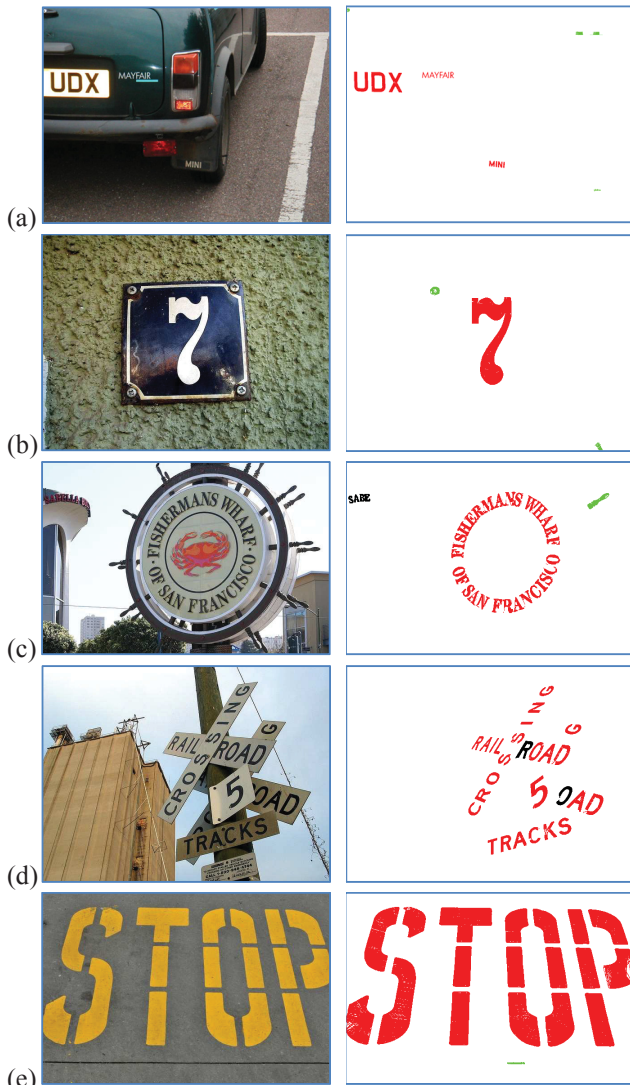


Fig. 1: Five examples. Left column lists original images. Right column lists resulting images.

large area or background objects passing through characters would frustrate most edge-based methods. Recently proposed Ring Radius Transform (RRT) [15] could be a remedy approach. One of our further works is to solve this common problem by extending LHA with the cooperation of RRT.

IV. CONCLUSION

An edge-ray filter which fully utilizes the spatial layout of edges is presented in this paper. Edges are detected by a combination of Canny and EPSF. The EQCA is employed to unify the complex edges and contour of broken character, which indeed increases the filtering strength as demonstrated in the second comparative study. Four regulations of LHA have the remarkable ability to filter out non-character edges and redundant rays. Benefiting from the absence of the assumption of straight text line, the proposed method can handle special scenarios like single character, curve text line and more complex layouts as shown in Fig.1. Moreover, the edge-ray filter practically provides the accurate character segmentation masks which can well gear towards OCR. Particularly, the method is applicable to the scene where both dark-on-bright and bright-on-dark characters appear simultaneously. The quantitative results demonstrate that the proposed method

TABLE II: Evolution of the number of (quasi-)connected edges.

| GECA | QECA | LHA | AFR_{QECA} | AFR_{LHA} |
|--------|-------|------|--------------|-------------|
| 352292 | 53618 | 9520 | 76.99% | 75.32% |

GECA = General Edge Connectivity Analysis; AFR = Average Filtering Ratio

achieves the comparable performance in terms of the top recall and acceptable precision. Beside integrating with RRT, another further work aims to increase the precision by introducing more information like color, extremal region and so on.

REFERENCES

- [1] K. Jung, K. Kim and A. Jain, Text Information Extraction in Images and Video: A Survey, *Pattern Recognition*, vol.37, no.5, pp.977-997, 2004.
- [2] D. Chen, J. Odobez and H. Bourlard, Text Detection and Recognition in Images and Video Frames, *Pattern Recognition*, vol.37, no.3, pp.595-608, 2004.
- [3] K. Subramanian, P. Natarajan, M. Decerbo, etc., Character-Stroke Detection for Text-Localization and Extraction, *ICDAR*, 2007.
- [4] V. Dinh, S. Chun and S. Cha, An Efficient Method for Text Detection in Video Based on Stroke Width Similarity, *ACCV*, 2007.
- [5] J. Zhang and R. Kasturi, Character Energy and Link Energy-Based Text Extraction in Scene Images, *ACCV*, 2010.
- [6] P. Shivakumara, W. Huang, T. Phan, etc., Accurate Video Text Detection through Classification of Low and High Contrast Images, *Pattern Recognition*, vol.43, no.6, pp.2165-2185, 2010.
- [7] P. Shivakumara, T. Phan and C. Tan, A Laplacian Approach to Multi-Oriented Text Detection in Video, *PAMI*, vol.33, no.2, pp.412-419, 2011.
- [8] L. Neumann and J. Matas, Real-Time Scene Text Localization and Recognition, *CVPR*, 2012.
- [9] B. Epshtein, E. Ofek, and Y. Wexler, Detecting Text in Natural Scenes with Stroke Width Transform, *CVPR*, 2010.
- [10] C. Yao, X. Bai, W. Liu, etc., Detecting Texts of Arbitrary Orientations in Natural Images, *CVPR*, 2012.
- [11] C. Shi, C. Wang, B. Xiao, etc., Scene Text Detection Using Graph Model Built Upon Maximally Stable Extremal Regions, *Pattern Recognition Letters*, vol.34, no.2, pp. 107-116, 2013.
- [12] C. Yi and Y. Tian, Text Extraction from Scene Images by Character Appearance and Structure Modeling, *Computer Vision and Image Understanding*, vol.117, no.2, pp.182-194, 2013.
- [13] N. Nikolaou, N. Papamarkos, Color Reduction for Complex Document Images, *International Journal of Imaging Systems and Technology*, vol.19, no.1, pp.14-26, 2009.
- [14] A. Shahab, F. Shafait and A. Dengel, ICDAR 2011 Robust Reading Competition Challenge 2: Reading Text in Scene Image, *ICDAR*, 2011.
- [15] P. Shivakumara, T. Phan, S. Bhowmick, etc., A Novel Ring Radius Transform for Video Character Reconstruction, *Pattern Recognition*, vol.46, no.1, pp.131-140, 2013.