Globally Optimal Text Line Extraction based on K-Shortest Paths algorithm

Liuan Wang, Wei Fan, Jun Sun
Fujitsu Research & Development Center CO., LTD
Beijing, China
{liuan.wang, fanwei, sunjun}@cn.fujitsu.com

Seiichi Uchida
Kyushu University
Fukuoka, Japan
uchida@ait.kyushu-u.ac.jp

***Abstract*—The task of text line extraction in images is a crucial prerequisite for content-based image understanding application. In this paper, we propose a novel text line extraction method based on k-shortest paths global optimization in images. Firstly, the candidate connected components are extracted by reformulating it as Maximal Stable Extremal Region (MSER) results in images. Then, the directed graph is built upon connected component nodes with edges comprising of unary and pairwise cost function. Finally, the text line extraction problem is solved using the k-shortest paths optimization algorithm by taking advantage of the particular structure of the directed graph. Experimental results on public dataset demonstrate the effectiveness of proposed method in comparison with state-of-the-art methods.**

Keywords—text line extraction; directed graph; cost function; k-shortest paths optimization;

I. INTRODUCTION

With the rapid increasing using of image captured devices, fast and accurate content-based image understanding systems have been received increasing attention due to its semantic information in recent years. As a crucial prerequisite, the text line extraction plays an important role in Optical Character Recognition applications. However, accurate and fast text line extraction is a challenging task due to large diversity of text size, font, orientation, noises and complex background. Some of challenging images from International Conference on Document Analysis and Recognition (ICDAR) competition dataset are shown in Fig.1.

Currently, a large number of state-of-the-art text line extraction methods have been proposed, which can be roughly classified into two categories [1]: region-based and connected components (CC)-based method. Region-based methods try to extract discriminative hand-crafted features in sliding windows and classify the local windows to text and non-text regions by well-trained classifiers. Hanif et al. [3] proposed a complete text localization boosting framework integrating feature and weak classifier selection based on computational complexity to construct the text detectors, and a neural network to learn the necessary rules for localization. Minetto et al. [4] describe a robust and accurate multi-resolution approach to detect and classify text regions in scenarios, the segmented regions are filtered out using shape-based classification, and neighboring characters are merged to generate text hypotheses. Wang et al. [5] perform multi-scale character detection via sliding window classification, which implements the features consisting of applying randomly chosen thresholds on randomly chosen entries in a HOG descriptor computed at the window location.

As opposed to region-based methods, the CC-based methods attempt to extract connected components from images, and then group them to text lines. Shi et al [6] formulated the text detection as a bi-label (text and non-text regions) segmentation problem, It used a graph model built upon MSERs to incorporate various text information sources into one framework, and the cost function could be optimally minimized via graph cut algorithm to get the final MSERs labeling results. Wang et al [7] proposed a generic text line extraction method, which can be applied on large categories of multi-orientated document images. In its coarse step, the text lines are generated from hierarchical edges reconstruction and cut by local linearity in the MSER spanning tree, in refinement step, the cut multi-components are re-connected based on the text line energy minimization in terms of text line consistency and fitting error. Yin et al [8] extracted MSERs as character candidates using minimizing regularized variations, and then the candidates are grouped into text lines by the single-link clustering algorithm.

Fig.1. Example of ICDAR competition image.

The combining methods may benefit from the advantage of both region-based and CC-based method. Huang et al [9] claimed a novel framework to tackle the low-quality texts by taking advantage of MSERs and sliding window based method. The MSERs operator can dramatically reduce the number of windows scanned and the sliding window with convolutional neural network (CNN) is applied to correctly separate the connections of multiple characters in components.

In this paper, we propose a novel text line extraction method based on global k-shortest paths optimization. The text lines are extracted by a global optimization strategy instead of extracting individual text one-by-one by geometric grouping. We can expect that it is beneficial to avoid inconsistent extraction results. Firstly, the MSERs are extracted as the candidate connected components, and then one directed graph is built upon the candidate connected components nodes with unary and pairwise cost function edges. Finally, we take advantage of the particular structure of the constructed directed graph and solve the text line extraction problem by k-shortest paths optimization algorithm. The flowchart of the proposed method is shown in Fig.2.

The main novel contributions of this paper are described as follows.

- The directed graph designed from candidate connected component nodes can effectively decrease the computational complexity for paths optimization in the graph model.
- We demonstrate that the text line extraction problem can be solved by the k-shortest paths optimization algorithm in the built directed graph with cost function of intrinsic consistency of the same text line.

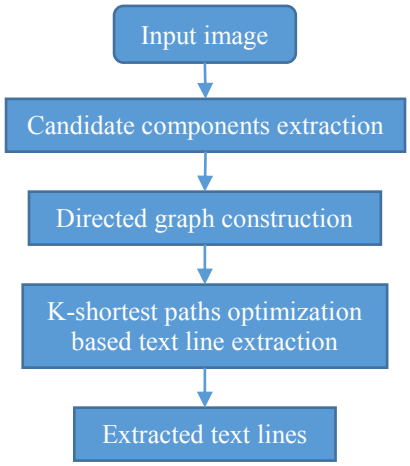

Fig.2. Flowchart of the proposed method.

The proceeding of this paper is organized as following. In section II, the candidate connected components extraction by MSER is described. Section III presents the construction of directed graph from the candidate connected components and cost function generation for graph edges. Section IV gives the text line extraction by k-shortest paths optimization algorithm. Finally, the experimental results of the proposed method are demonstrated in section V, and the conclusions are drawn in sect VI.

II. CANDIDATE CONNECTED COMPONENTS EXTRACTION

A. Candiate Character Extraction

The Maximal Stable Extremal Region [10] has been considered as one of the best candidate connected components extraction method, which has won the first place in ICDAR2011 [13] and ICDAR2013 competitions [14] and achieved promising performance. An Extremal Region (ER) is well defined as a connected component region on image if all the set of pixels have larger or lower intensity values than its outer boundary intensity. One Extremal Region is maximally stable when it has a minimum, and the Maximal Stable Extremal Region can be regarded as one virtually unchanged local binarization over a large range of thresholds. The MSER method is very efficient for multi-scale detection and low quality connected components with near linear complexity [11]. Two polarities of connected components are extracted based on the MSER method: black MSERs in white background and white MSERs in black background.

Although the MSER method can achieve promising recall and extract most of the candidate connected components even in low quality images, it suffers from the noises in terms of low precision. Thanks to the MSER structure of rooted tree, the duplicate MSERs can be efficiently pruned by searching all the MSER sequences. Similar MSER pruning method [8] are used to discard the repeating noises MSER connected components.

B. Noise Candidate Removal

The pruning algorithm can effectively remove repeating MSERs. However, any non-text objects, such as bikes, leaves, and so on, will generate noise MSER connected component regions. One classifiers fusion strategy [7] between boosting and Convolutional Neural Network is adopted to filtered out the noise candidate connected components. Firstly, the fast boosting classifier with hand crafted geometric features is utilized to classify the candidate MSERs to text and non-text. The non-text MSERs are discarded as noises, it can effectively remove lot of noise MSERs with small computational expense. The text candidate with high recognition certainty are retained as candidates, and we re-recognize the text with low recognition probability by CNN, which can learn high-level features to identify the text MSERs from the non-text outliers robustly. Finally, the negative noises are removed, and positive texts are saved.

III. DIRECTED GRAPH CONSTRUCTION

A directed graph that connects all neighboring MSER pairs is created. Correct text lines are assumed included as paths in the directed graph. Then, each edge between two neighboring MSERs is attached an appropriate cost in the directed graph. We can have a globally optimal text line extraction results by extracting the paths which give the minimum total cost.

A. Constraints of potential directed graph edges

There are two constraints for the potential directed graph edges, and two candidate MSER components are potential linking when the criterions are satisfied. One is the distance constraint as following to avoid unnecessary edges.

$$dist(m_i, m_j) < k * \min(\max(w_i, h_i), \max(w_j, h_j)) \quad (1)$$

where $dist(\cdot,\cdot)$ is the distance of two centers of MSERs, w_i, h_i and w_j, h_j denote the width and height of the bounding boxes of the i^{th} and j^{th} MSERs, respectively.

The other criterion is overlapping constraint in the assigned direction from left to right, it's reasonable for the arrangement of the text lines. Two MSER components satisfying these two constraints are potential linking in text line.

B. Directed Graph Construction

To reduce the computational complexity of the k-shortest paths optimization, a directed graph $G = (V, E)$ is constructed composed of vertices V from MSER connected components and directed edges E connecting these vertices. Supposed that each MSER component center is a vertex v_i, there is one edge $e_{i,j}$ between two linking vertices v_i and v_j in the directed graph. The vertices and edges of the directed graph are $V = \{v_i\}_{i=1}^{n}$ and $E = \{e_{i,j} \mid v_i, v_j \in V\}$, where denotes the number of MSER components vertex. The vertices of the directed graph are location of MSER components, each vertex can be the start or end location of one text line. Therefore, two additional virtual vertices v_{source} (green dash circle) and v_{sink} (blue dash circle) are added to our directed graph, where v_{source} is the possible graph source point, and v_{sink} denotes the text line sink point. There is one edge from each MSER component vertex to the source and sink vertex, and the cost value of these edges are set as 0 to allow each MSER component to be the start or end location of the text line at no cost.

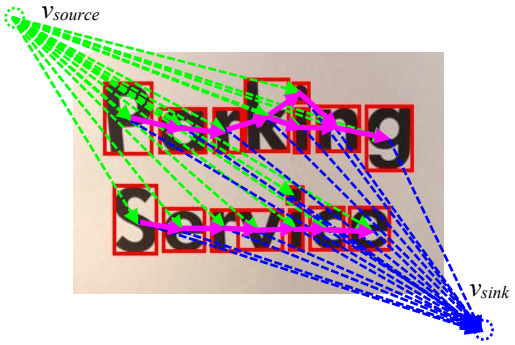

Fig.3. Example of directed graph.

The directed graph edges are built in assigned direction from left to right according to the position relationship of the MSER vertices when potential linking existing. For each vertex, we firstly find out the nearest MSER component vertex v_j from the set of potential linking MSER components $\{v_i\}_{i=1}^{m}$, and then collect all the MSER components, which exists overlapping in vertical direction. Finally, the edges are built from the current vertex to the select MSER components. Fig.3. gives one example of the built directed graph, the green dash lines, the blue dash lines and purple dash lines denote the edges between MSER vertex and graph source, sink point and linking MSER vertex.

C. Cost Function of Graph Edges

The edge cost function $c(e_{i,j})$ measures the cost value from one MSER vertex to the other MSER vertex in the directed graph edge. Unary cost function and pairwise cost function are employed to calculate the cost of each edge. Note that there is no cost for edges between each MSER vertex and virtual source point, as the same as sink vertex.

1) Unary cost function: Unary cost function measures the cost value for classifying the candidate MSER vertex into text.

a) Probability of recognition engine: The probability of recogntion engine is a strong feature for discriminating the text from noises.

b) Variation of MSER: The text MSER component is virtually unchanged over a large range of thresholds to distinguish it to the background, the text MSERs tend to have smaller variation.

c) Occupation ratio: Text MSER componects often share common occupation ratio, too large or too small of which will lead to noises.

2) Pairwise cost function: Pairwise cost function measures the cost value for the discontinuity of two linking text MSER candidate in the text line.

a) Location distacne of adjacent vertices: The text line is composed of some individual text in specific order. The distance of adjacent MSER components in text line is small, while there is no regular pattern for noises.

b) Overlapping of adjacents vertices: All the components in the text line are arranged in one straight line, and there is overlaping between adjacent text components.

c) Color similarity: Adjacent text vertices in the same text line shares similarity of color characteristic.

IV. K-Shortest Paths Optimization

One text line can be considered as one path flow from the source text to the sink text in the constrict direction, therefore, the text line extraction problem can be reformulated as a global path flow optimization results. Fig.4 gives one example of the paths optimization based text line extraction, the purple MSER components path flows are the route of text line.

Fig. 4. Paths optimization based text line extraction.

Our text line extraction problem can be treated as a minimum cost flow problem with a 0-1 flow constraint. If 1, it suggests the edge is part of a text line. The k-shortest paths algorithm is well-studied and widely applied [16] for path selecting and routing in the directed graph. We take advantage of the particular structure of constructed directed graph to obtain the global optimal text line solution by the k-shortest paths optimization algorithm.

The target text line extraction problem is to find the optimal solution that minimizes the cost function between the source vertex v_{source} and sink vertex v_{sink} in the directed graph by (2).

$$f^* = \arg\min_{G=(V,E)} \sum c(e_{i,j}) \cdot l(e_{i,j}) \tag{2}$$

where $c(e_{i,j})$ and $l(e_{i,j})$ denote the cost function and label of edges $e_{i,j}$ for vertex i, j respectively.

The k-shortest paths optimization algorithm find k paths $\{p_1, p_2, \cdots, p_k\}$ iteratively with minimum total cost value in the directed graph, where k is settled. Any path between v_{source} and v_{sink} in the directed graph represents a feasible path flow of text line. In our text line extraction case, no text MSER component cannot be shared by two text lines, which means the extracted k paths should be vertex disjoint, and each vertex should be included in one path at most.

In initialization, Dijkstra algorithm [17] is used to generate the single shortest path, At the n^{th} iteration, supposed n-shortest text line paths $P_n = \{p_1, p_2, \cdots, p_n\}$ with minimum total cost are found using the previous $(n\text{-}1)$ shortest text line paths.

The single shortest text line path cost value of p_l can be calculated by all of the graph edges $e_{i,j}$ belonging to the l^{th} shortest path.

$$c(p_l) = \sum_{e_{i,j} \in p_l} c(e_{i,j}) \tag{3}$$

The total cost value of the n-shortest paths at n^{th} iteration is formulated by (4).

$$c(P_n) = \sum_{i=1}^{n} c(p_i) \tag{4}$$

We compare the total cost value $c(P_{n+1})$ of new iteration $(n+1)$ with the cost of previous iteration $c(P_n)$, the total path costs are monotonically increasing, and the global minimum is achieved when the cost value change sign and become decreasing as show in (5). Meanwhile, the optimal parameter k is obtained. Fig. 5 gives one example of text line extraction results based on the k-shortest paths optimization method, and the purple path flows are text lines.

$$\begin{cases} c(P_{n-1}) \le c(P_n) \\ c(P_n) \ge c(P_{n+1}) \end{cases} \tag{5}$$

Fig.5. Example of text line extraction results by k-shortest paths optimization.

V. EXPERIMENTAL RESULTS

To evaluate the effectiveness of the proposed text line extraction method based on k-shortest paths optimization algorithm, we demonstrate our experimental results on public benchmark scene text detection dataset ICDAR 2003 [18], and the traditional performance metrics, precision, recall and f-measure are adopted to quantify the performance.

The performance evaluation method is to compare the region matching ratio $mr(i, j)$ between extracted text lines region and the ground truth (GT) text lines region.

$$mr(i, j) = \frac{f(E_i \cap G_j \cap I)}{f((E_i \cup G_j) \cap I)} \tag{6}$$

where I is the image region, E_i and G_j denote the i^{th} extracted text line region and the j^{th} ground truth region, respectively. $f(\cdot)$ is a function computing the intersection region.

The precision rate and recall rate are defined based on the text line area matching ratio.

$$precision = \frac{\#correct}{\#correct + \#false}, recall = \frac{\#correct}{\#GT} \tag{7}$$

Another f-measure is widely used combining the recall rate and precision rate with parameter α, which is set 0.5 for equal weight.

$$f = \frac{1}{\alpha \cdot recall + (1-\alpha) \cdot precision} \tag{8}$$

The experimental results of k-shortest paths optimization based text line extraction are illustrated in table.1 by comparing with the state-of-the-art method. In ICDAR 2003 Dataset, we achieve the precision of 0.66, recall of 0.62, and f-measure of 0.64. Our k-shortest paths optimization method is mainly affected and sensitive to the false positive MSER components, our further plan is to make the k-shortest paths optimization algorithm robust to false positive components. Fig.6 demonstrates some successful and failure examples of extracted text lines in ICDAR 2003 dataset. Green, red, blue text line bounding boxes represent the correct, false positive, and failure of text line extraction, respectively.

TABLE I. EXPERIMENTAL RESULTS OF K-SHORTEST PATHS OPTIMIZATION BASED TEXT LINE EXTRACTION ON ICDAR2003 DATASET

	precision	***recall***	***F-measure***
Koo [19]	0.78	0.65	0.71
Our method	**0.66**	**0.62**	**0.64**
Becker [20]	0.62	0.67	0.64
Li [21]	0.62	0.65	0.63
Neumann [22]	0.59	0.55	0.57
Zhou [23]	0.57	0.50	0.53

Fig.6. Example of successful and failure text line extraction.

VI. CONCLUSION

In this paper, we propose a novel efficient text line extraction method based on the k-shortest paths optimization algorithm. Firstly, Maximal Stable Extremal Regions are extracted as the candidate connected components, and then we build a directed graph rather than traditional undirected graph by setting the candidate MSER components as the graph vertex, and the cost value of each edge between two connecting vertices is calculated by unary cost function and pairwise cost function. Finally, the text lines are extracted by k-shortest paths optimization method in the constructed directed graph. The major contributions are the directed graph construction with two virtual source and sink vertex, and formulating the text line extraction problem by k-shortest paths optimization algorithm. The directed graph can effectively decrease the computational complexity for paths optimization in the graph model in comparison with undirected graph. Our further works will focus on robust k-shortest paths optimization with false positive components and multilingual multi-oriented text line extraction.

REFERENCES

[1] K. Jung, K. I. Kim, and A. K. Jain, "Text Information Extraction in Images and Video: A Survey," Pattern Recognition, 37, pp. 977-997, 2004.

[2] Q. X. Ye, D. Doermann, "Text Detection and Recognition in Imagery: A Survey," IEEE Trans. Pattern Analysis and Machine Intellligence, Vol. 37, no.7, pp. 1480-1500, 2014.

[3] S. M. Hanif, L. Prevost, "Text Detection and Localization in Complex Scene Images using Constrained Adaboost Algorithm," 10th international conference on Document Analysis and Recognition, pp. 1-5, 2009.

[4] R. Minetto, N. Thome, M. Cord, J. Fabrizio, B. Marcotegui, "Snoopertext: A multiresolution system for text detection in complex visual scenes" 17th International Conference on Image Processing, pp. 3861-3864, 2010.

[5] K. Wang, B. Babenko, S. Belongie, "End-to-End Scene Text Recognition", 2011 IEEE International Conference on Computer Vision, pp. 1457-1464, 2011.

[6] C. Z. Shi, C. H. Wang, B. H. Xiao, Y. Zhang, S. Gao, " Scene text detection using graph model built upon maximally stable exremal regions" Pattern Reconition Letters, 34, pp. 107-116, 2013.

[7] L. A. Wang, W. Fan, J. Sun, S. Naoi, T. Hiroshi, "Text Line Extraction in Document Images", 13th International Conference on Document Analysis and Recognition, pp. 191-195, 2015.

[8] X. C. Yin, X. W. Yin, K. Z. Huang, H. W. Gao, "Robust Text Detecion in Nautral Scene Images," IEEE Transactions on Pattern Analysis and Machine Interlligence, vol 36, issue 5, pp. 970-983, 2013.

[9] W. L. Huang, Y. Qiao, X. O. Tang, "Robust Scene Text Detection with Convolution Neural Networks Induced MSER Trees," 13th European conference on Computer Vision, pp. 497-511, 2014.

[10] J. Matas, O. Chum, M. Urban, T. Pajdla, "Robust Wide Baseline Stereo from Maximally Stable Extremal Regions," British Machine Vision Computing, vol 22, issue 10, pp. 961-967, 2002.

[11] D. Nister, H. Stewenius, "Linear Time Maximally Stable Extremal Regions," 10th European conference on Computer Vision, pp. 183-196, 2008.

[12] L. Sun, Q. Huo, W. Jia, K. Chen, "Robust Text Detection in Natural Scene Images by Generalized Color-enhanced Contrasting Extremal Region and Neural Networks," 22nd International Conference on Pattern Recognition, pp. 2715-2720, 2014.

[13] A. Shahab, F. Shafait, A Dengel, "ICDAR 2011 Robust Reading Competition Challenge 2: Reading Text in Scene Images," International Conference on Document Analysis and Recognition, pp. 1491-1496, 2011.

[14] D. Karatzas, F. Shafait, S. Uchida et al. "ICDAR 2013 Robust Reading Competition," 12th International Conference on Document Analysis and Recognition, pp. 1484-1493, 2013.

[15] J. W. Suurballe, "Disjoint Paths in a Network," Networks, vol.4, pp. 125-145, 1974.

[16] J. Berclaz, F. Fleuret, E. Turetken, P. Fua, "Multiple Object Tracking using K-Shortest Paths Optimization," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 33, issue 9, pp. 1806-1819, 2011.

[17] E. W. Dijkstra, "A Note on Two Problems in connexion with graphs," Numerische Mathematik, vol. 1, pp. 269-271, 1959.

[18] S. M. Lucas, A. Panaretos, L. Sosa, "ICDAR 2003 Robust Reading Competition: Entries, Results, and Feature Directions," International Journal on Document Analysis and Recognition, vol. 7, pp. 105-122, 2005.

[19] H. Koo, D. Kim, "Scene Text Detection via Connected Component Clustering and Non-text Filtering," IEEE Transaction on Image Processing, vol. 22, issue 6, pp. 2296-2305, 2013.

[20] S. M. Lucus, "Text Locating Competition Results," IEEE International Conference on Document Analysis and Recognition, pp. 80-85, 2005.

[21] Y. Li, C. H. Shen, W. J. Jia, A. Hengel, "Leveraging Surrounding Context for Scene Text Detection," 20th IEEE International Conference on Image Processing, pp. 2264-2268, 2013.

[22] L. Neumann, J. Matas, "A Method for Text Localization and Recognition in Real-World Images," 10th Asian Conference on Computer Vision, pp. 770-783, 2010.

[23] G. zhou, Y. Liu, Z. Q. Tian, Y. Q. Su, "A New Hybrid Method to Detect Text in Natural Scene," 18th IEEE International Conference on Image Processing, pp. 2605-2608, 2011.

[24] C. Yao, X. Bai, W. Liu, Y. Ma, Z. W. Tu, "Detecting Texts of Arbitrary Orientations in Natural Images," 2012 IEEE Conference on Computer Vision and Pattern Recogniton, vol. 8, pp. 1083-1090, 2012.

[25] L. Sun, Q. Huo, W. Jia, K. Chen, "Robust Text Detection in Natural Scene Images by Generalized Color-enhanced Contrasting Extremal Region and Neural Networks," 22nd International Conference on Pattern, pp. 2715-2720, 2014.

[26] C. Shi, B. Xiao, C. Wang, Y. Zhang, "Graph-based Background Suppression For Scene Text Detection," 10th IAPR International Workshop on Document Analysis Systems, pp. 210-214, 2012.