Text Location Algorithm Based on Graph-Cut Model with Unary and Binary Features

Fengqin Yu[1] , Yaya Liu[2]
School of Internet of Things Engineering, Jiangnan University, Wuxi, China
1: e-mail: yufq@jiangnan.edu.cn; 2: e-mail: 8489672@163.com

***Abstract*—In this paper, we propose a location algorithm of the text regions in an image based on a graph cut model with unary and binary features. First, the most stable extremal regions are detected as candidate regions. Then, we define the energy function using unary and binary features of the candidate regions. The text classification is obtained by the optimal segmentation. Lastly, the final positioning is obtained using text aggregation. The simulation results show that the proposed algorithm is better than several classic algorithms in terms of precision rate and standard measurement and can precisely locate text regions in an image.**

Keywords- text location; graph-cut model; unary text features; binary text features.

I. INTRODUCTION

Nowadays, images have become an important carrier of information, while the text embedded in images provides semantic information. Text localization is the positioning of the text area in an image and has become an important research topic in image interpretation and pattern recognition. Text location methods can be divided into three main kinds, namely: edge-based [1]-[2], connected-component-based [3]-[5] and texture-based [6]-[8].

We propose a text location algorithm based on the multi-feature graph-cut model described in Pan *et al.* [3]. First, candidate text regions are identified from the Maximally Stable Extremal Regions (MSER) in a contrast-enhanced input image. Each candidate region is treated as a vertex in the graph-cut model. We define an energy function containing a regional term and a boundary term based on the unary features and the binary features of the candidate regions. Subsequently, the energy function is minimized to classify the candidate text regions and remove the background. Finally, we connect adjacent text regions (text aggregation) and quantify their locations.

The rest of the paper is structured as follows. In Section 2, we present the graph-cut model with unary and binary features. In Section 3, we outline the algorithm implementation steps. Section 4 presents the simulation experiment and results analysis. We conclude the paper in Section 5.

II. GRAGH-CUT MODEL WITH UNARY AND BINARY FEATURES

A. Gragh-Cut Model

The graph-cut model provides a means of image segmentation by mapping an image into a weighted graph, where a chosen pixel in the original image represents the vertex of the graph and the relation between this pixel and the domain translates into the edge of the graph. The energy function is formed with the edge weights, and a solution that minimizes the energy function represents an optimal way of partitioning the given image into text regions and non-text regions (background).

In the graph-cut model, the pixels of the input image correspond to the vertices in the output graph one to one. Every edge has a weight. According to the edge weights, the energy function is formulated as follows:

$$E(L)=\sum_{p\in V}R_p\left(l_p\right)+\lambda\sum_{\{p,q\}\in E}B_{\{p,q\}}*\delta\left(l_p,l_q\right) \tag{1}$$

where, $L=(l_1,l_2,\ldots l_n)$ are two value vectors, n is the number of vertices, L is the label vector of the graph, and each label vector corresponds to the cut set of an image. V and E are the set of vertices and edges, respectively. p and q are regions. $\sum_{p\in V}R_p\left(l_p\right)$ is the region term, $\sum_{\{p,q\}\in N}B_{\{p,q\}}$ is the boundary term, λ is the weight factor, and $\delta(l_p,l_q)$ is the Dirac function.

B. Unary text feature

A unary text feature describes the characteristic of the text region–basically its probability of being text, as opposed to being background. The unary text feature constitutes the regional term in the energy function. The method by Pan *et al.* [3] cannot comprehensively represent text characteristics because it employs only prior characteristic rules, such as standard width and height, aspect ratio, and occupancy. In comparison, our unary text feature includes the edge gradient, center surround histogram, and stroke width coefficient of variation.

C. Binary text feature

A binary text feature describes the relationship between a region of concern and its neighborhood. It reflects the probability of the two regions being the same or different types. The more similar the features of the two regions are, the higher the probability that the two regions are of the same type. The binary text feature constitutes the boundary term in the energy function. Pan *et al.* [3] include only the regional features as the binary text feature; whereas, we consider color images, where the color distribution and the regional similarity are included as binary text features. Two regions p and q, are considered as being adjacent if they satisfy the following criterion:

 ISBN: 978-1-61208-570-8

$$dis(p,q) < 2 \times \min\left[\max\left(w_p, h_p\right), \max(w_q + h_q)\right] \quad (2)$$

where, w and h are the width and height of the region and $dis(p,q)$ is the Euclidean distance between the centroids of the two regions p and q.

D. Energy function with Unary and Binary Features

The regional term of the energy function is the sum of the edge weights, and the edge is defined as the link between the vertex and the endpoint; it reflects the regional characteristic. Since the unary text features, edge gradient, center surround histogram, and stroke width coefficient of variation well describe the regional characteristic, we use these features to establish the regional term for the region p .

III. ALGORITHM IMPLEMENTATION STEPS

Step1. The contrast of the input image is enhanced and the MSERs are detected as candidate text regions. The input image is divided into a light-dark image via contrast enhancement.

Step2. Candidate text regions are filtered based on heuristic rules. Each region is treated as a vertex, and a graph model is constructed.

Step3. The regional term of the energy function is formulated from the unary text features, the edge gradient, the center surround histogram, and the stroke width coefficient of variation of the candidate regions.

Step4. The boundary term of the energy function is formulated from the binary text features, the color distribution, and the regional similarity extracted from the candidate regions.

Step5. An optimal segmentation of the candidate regions is obtained by minimizing the energy function, with a weighting factor λ=0.5. The fore-ground are text regions and the rest are removed.

Step6. Adjacent characters are connected based on text aggregation. The bright-dark text image is combined with the original image to produce the final text positioning result.

IV. SIMULATION EXPERIMENT AND RESULTS ANALYSIS

In order to test and verify the validity of our algorithm in segmenting and labeling text regions, we employed the dataset publicly available from International Conference on Document Analysis and Recognition (ICDAR), which is composed mainly of various indoor and outdoor images taken with a digital camera.

TABLE I. PERFORMANCE COMPARISON

Method	Precision Rate	Recall rate	Standard Measure
Proposed Algorithm	0.75	0.68	0.72
Pan [3]	0.67	0.70	0.69
Rodrige [7]	0.74	0.63	0.68
Yi [4]	0.71	0,62	0.67
Epstein [2]	0.73	0.60	0.66

Table 1 compares the performance of the proposed algorithm against several classic algorithms. Compared with [2][3][4][7], the precision rate and the standard measure of the proposed algorithm increased, but the recall rate decreased.

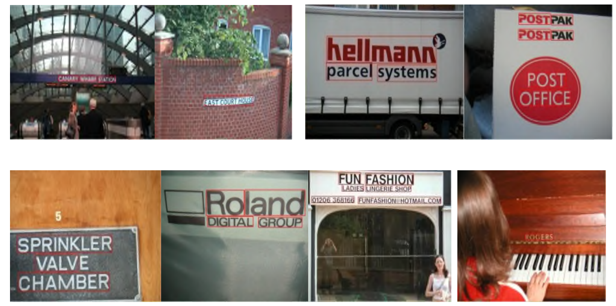

Figure 1. Some location results

As demonstrated in Figure 1, the proposed algorithm successfully isolated text regions from complex backgrounds in various test natural scene images.

V. CONCLUSION

We propose a text localization algorithm that is based on the multi-feature graph-cut model and achieves text classification through optimal segmentation. First, the MSERs are extracted as candidate text regions. Subsequently, each region is treated as a vertex and the graph-cut model is established. An energy function with the regional and boundary terms is constructed from the unary and binary text features, and candidate regions are classified by minimizing the energy function. Finally, adjacent characters are connected together based on text aggregation. We demonstrate that the proposed algorithm is able to locate precisely text regions in a variety of natural scene images.

REFERENCES

[1] B. Epshtein, E. Ofek,and Y. Wexler, “Detecting text in natural scenes with stroke width transform, ”Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.2963-2970, 2010.

[2] Z. Zhou and C. L. Tan, “Edge based binarization for video text images. ”Proceeding of 20th International Conference on Pattern Recognition. Istanbul,Turkey, pp.133-136,2010.

[3] Y. Pan, X. Hou, and C. Liu,“A hybrid approach to detect and localize texts in natural scene images,” IEEE Trans.on Image Processing, vol. 20, pp.800-813, 2011.

[4] C. C. Yi and Y. L. Tian, “Text string detection from natural scenes by structure-based partition and grouping,”IEEE Trans.on Image Processing, vol. 20, pp. 2594-2605, 2011.

[5] X.Yin, K. Huang, and H. Hao, “Robust text detection in natural scene images,”IEEE Trans.on Pattern Analysis and Machine Intelligence, vol.36, pp.70–983,2013.

[6] J.Ye, L. L. Huang, and X. L. Hao, “Neural network based text detection in videos using local binary patterns,”IEEE Chinese Conference on Pattern Recognition(CCPR), pp.1-5, 2009

[7] R.Minetto, N. Thome, and M. Cord, “SnooperText: A text detection system for automatic indexing of urban scenes,” Computer Vision and Image Understanding, Vol.122, pp.92-104, 2014.

[8] S.Kumar, R. Gupta, and N.Khanna, “Text extraction and document image segmentation using matched wavelets and MRF model,” IEEE Trans.on Image Processing, Vol. 16, pp. 2117-2128, 2007.