A Text Image Segmentation Algorithm Based on Spectral Clustering: A Survey

Dung Phan
Pattern Recognition Lab
School of Computer Engineering
Chonnam National University
Email: dungptmy@gmail.com

Abstract

The presented method is used to segment text in natural scene images. This project refers to [1] for implementing a text segmentation algorithm. The proposed algorithm uses the normalized graph cut as the measure for spectral clustering which examines the affinities (similarities) between groups of pixel that have the same grayscale level. That requires much smaller spatial costs and much lower computation complexity because size of affinity matrix depends on the number of gray-level. Experimental result show the superior performance of the proposed method compared to the typical thresholding algorithms.

Keywords — Text Segmentation, Graph cut, Normalized cut, Spectral Clustering

I. INTRODUCTION

Text segmentation is a precursor to text retrieval, automatic summarization, information retrieval (IR); language modeling (LM) and natural language processing (NLP). Images generally contain rich messages from textual information, such as street name, construction identification and a variety of signal boards, etc. The textual information assists in understanding the essential content of the images. For instance, street name, signal boards and construction identification in a natural environment can be captured into images by cameras and the textual information will be detected, segmented, and recognized automatically by smartphone. These messages then can be synchronized as human voice to be used as instructions for visually impaired person. In addition to the example, textual information extraction plays a major role in images retrieval based on contents, cars auto-drive, vehicle plate recognition and automatics.

Almost OCR (Optical Character System) engine can only deal with printed characters against clean backgrounds and cannot handle character embedded in shaded, textured or complex backgrounds. However, in natural scene, text can have different backgrounds and character in text can also have a lot of forms. Therefore, text detection is very necessary in order to remove the background. In Pan's (2007,pp.412-416), a simple global threshold achieved by using Otsu's (Otsu,1979,pp.62-66) thresholding technique. Lienhart (2002,pp.256-268) performed the binarization using the intensity value halfway between the intensity of the text colors and the background color as a threshold. Wu (1999,pp.1224-1229) proposed a simple histogram-based algorithm to automatically find the threshold value for each text region, making the text segmentation process more efficient. Due to its simplicity and efficiency, thresholding is a widely used method for solving this problem. But, it could not handle the cases when backgrounds have the similar color or intensity to that of the text strokes. Meanwhile, besides the changing backgrounds, texts are also changing slightly due to edge blur, image quality degrading due to video compression.

In this paper, they propose a new text segmentation method based on spectral clustering. In our approach, the histogram of intensity is used for the object of grouping. We partition the image into two parts using the gray levels of an image rather than the image pixels. For most images, the number of gray levels is much smaller than the number of pixels. Therefore, the proposed algorithm occupies much smaller storage space and requires much lower computational costs and implementation complexity than other similar algorithms.

II. BACKGROUND AND RELATED WORK

A. Theory of Spectral Graph

Given a weighted undirected graph $G = (V, E)$, where the nodes of graph are the points in the feature space and an edge is formed between every pair of nodes. The weight on each edge, $w(i, j)$, is a function of similarity between node i and j. A graph $G = (V, E)$ can be partitioned into two disjoint subsets A and B, subject to $A \cap B = \emptyset$ *and* $A \cup B = V$, by simply removing edges connecting the two parts.
The cut between two groups A and B is defined as the sum of all the weights being cut:

$$cut\ (A, B) = \sum_{i \in A, j \in B} w(i, j) \qquad (1)$$

The optimal partitioning of a graph is the one that minimized this cut value. Although there is an exponential number of such partition, finding the minimum cut of a graph is well-studied problem and there are efficient algorithms for solving it. *Normalized Cut*, is proposed by Shi [2], show a new measure of disassociation between two groups. Instead of looking at the value of total edge weight connecting the two partitions, this method measure computes the cut cost as the fraction of total edge connections to all the nodes in the graph.

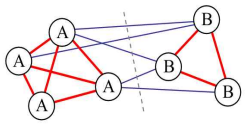
The disassociation measure is defined:

$$Nout\ (A,B) = \frac{out\ (A,B)}{assoc\ \ (A,V)} + \frac{out\ (A,B)}{assoc\ \ (B,V)} \qquad (2)$$

where $assoc\ \ (A,A) = \sum_{i\in A, j\in B} w(i,j)$ is the association (sum of all weights) within a cluster and $assoc\ \ (A,V) = assoc\ \ (A,A) + out\ (A,B)$ is the sum of all weights associated with nodes in A. Figure 1 shows how to cuts the associations. And now, the minimal *Ncut* value is just corresponding to the optimal bipartition of the graph. Unfortunately, computing the optimal normalized cut is NP-complete. Instead, Shi [2] suggest computing a real-valued assignment of nodes to groups. Shi [2] show that minimizing *Ncut* equivalent to solving the eigenvalue system:

$$(\boldsymbol{D} - \boldsymbol{W})y = \lambda \boldsymbol{D} y \qquad (3)$$

where D is diagonal matrix, $D_i = \sum_{j\in W} w(i,j)$, W is a symmetric matrix with size of NxN, λ is the eigenvalue and y is the eigenvector. Shi and Malik [2] have proved that the second smallest eigenvector of the eigensystem is the real value solution to the normalized cut problem.



(a)

	A	B	sum
A	$assoc(A,A)$	$cut(A,B)$	$assoc(A,V)$
B	$cut(B,A)$	$assoc(B,B)$	$assoc(B,V)$
sum	$assoc(A,V)$	$assoc(B,v)$	

(b)

Figure 1: Sample weighted graph and its normalized cut: (a) A small sample graph and its smallest normalized cut. (b) Tabular form of the association and cuts for this graph.

B. Proposed Method

Given an image, we can construct a histogram-based matrix M by computing the all weights of nodes in the corresponding graph. Let *M* be the *affinity matrix*, we can get a complete approach of image segmentation using spectral clustering [2]. Suppose $V = \{(i,j): i = 0,1,..,n_h - 1; j = 0,1,..,n_2 - 1;\}$, $H = \{H_0, H_1, \dots, H_L\}, LL = \{0,1,\dots,L\}$ *where* n_h *and* n_w is the height and the width of the image, respectively. H represents the histogram of gray, $f(x,y)$ is the gray value of position (x,y). Then, V,H and $f(x,y)$ satisfy the following formulas

$$(x,y) \in H_l, l \in \{0,1,\dots,L\}, \quad \forall (x,y) \in V \qquad (4)$$

$$H_l = \{(x,y): f(x,y) = l, (x,y) \in V\}, l \in LL \qquad (5)$$

$$\bigcup_{l=0}^{L} H_l = V, H_i \cap H_j = \emptyset, i \neq j, i,j \in LL \qquad (6)$$

Using just the intensity value of the pixels and their spatial location, we can define the graph edge weight connection the two nodes i and j as:

$$w_{ij} = e^{-\frac{\|F(i)-F(j)\|_2{}^2}{\sigma_l}} * \begin{cases} e^{-\frac{\|X(i)-X(j)\|_2{}^2}{\sigma_x}}, if\ \ \|X(i)-X(j)\| < r \\ 0 \qquad\qquad\qquad\qquad .otherwise \end{cases} \qquad (7)$$

where $F(i)$ is a feature vector base on intensity of node I and $X(i)$ is the spatial location at that node, σ_l *and* σ_x are scale factors used to adjust the variation of gray or spatial location between nodes, r is used to decide the number of nodes from node i to j.

And then, we can get a bipartition $V = \{A,B\}$ corresponding to the graph $G = \{V,E\}$, where $A = \cup_{k\in L_A} H_k, B = \cup_{k\in L_B} H_k$ *and* $L_A \cap L_B = \emptyset, L_A \cup L_B = LL$. Let $out\ (H_i,H_j) = \sum_{i\in H_i, j\in H_j} w(i,j)$ be the total connection weights from nodes in H_i with gray level I to all nodes in H_j with gray level j, we can rewrite the above formulas as:

$$out\ (A,B) = \sum_{i\in L_A}\sum_{j\in L_B} out\ (H_i,H_j) \qquad (8)$$

$$asso\ \ (A,A) = \sum_{i\in L_A}\sum_{j\in L_A} out\ (H_i,H_j) \qquad (9)$$

$$asso\ \ (B,B) = \sum_{i\in L_B}\sum_{j\in L_B} out\ (H_i,H_j) \qquad (10)$$

Since
$assoc\ \ (A,V) = assoc\ \ (A,A) + out\ (A,B)$,
$assoc\ \ (B,V) = assoc\ \ (B,B) + out\ (A,B)$, we can rewrite (2) as:

$$Nout\ (A,B) = \frac{out\ (A,B)}{assoc\ \ (A,A) + out\ (A,B)} + \frac{out\ (A,B)}{assoc\ \ (B,B) + out\ (A,B)} \qquad (11)$$

The histogram-based matrix M is defined:

$$\begin{cases} M\ is\ \ an\ \ LxL\ \ symmetrical \qquad matrix \\ m_{ij} = out\ (H_i,H_j) \\ m_{ij} = m_{ji} \end{cases}$$

Figure 3 show the workflow of this method.

III. EXPERIMENTAL RESULT

We perform test on scene text database which some normal text images and abnormal text image. In the following experiments the parameter setting in formula (7) are $\sigma_l = 50; \sigma_x = 5; r = 5; L = 100;$ maxCutValue $= 0.14$. For case 3,4 that are images with light affection, the gray scale factor should be $\sigma_l = 70$. This method is compared with Otsu method which is a simple but classic solution employed by many text segmentation schemes.

Database: http://ai.kaist.ac.kr/home/node/85

Figure 2 show the result of proposed method and Otsu's Method.

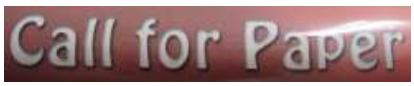
Test case	Input Image	Otsu's Result	Our Result
1			
2			
3			
4			
5			
6			
7			
8			

Figure 2: Text Image Segmentation Result:

a. **Normal cases, proposed method work well but Otsu's method doesn't**
b. **Blue rows are normal cases**

c. **Rred rows are abnormal case that both methods don't work well.**

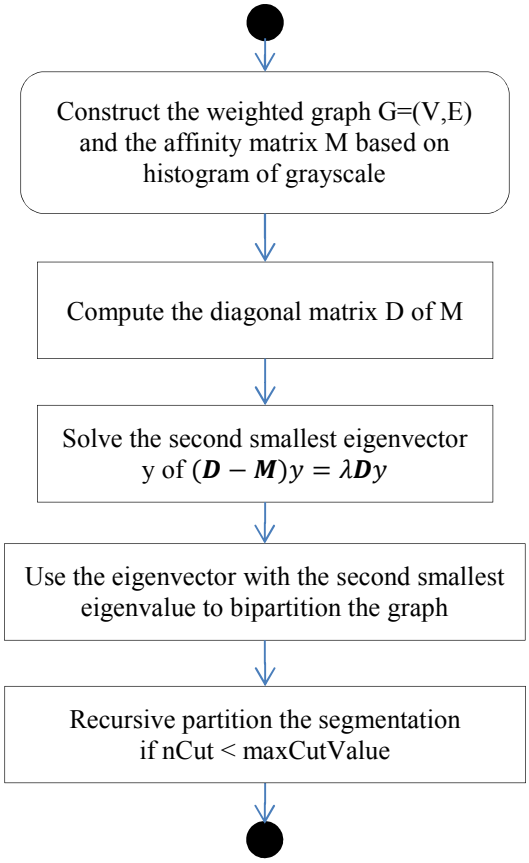


Figure 3: Flowchart of text segmentation based on Spectral Clustering

IV. CONCLUSION AND FUTURE WORK

Accurate retrieval of textual information from the image that exists in the real environment is critical to understand images. The key part of the research is to retrieve the characters from the text image area. Thresholding methods are very popular in text image segmentation. However, for the image with complex background, it cannot work well. Spectral clustering can resolve the issue by using spectral graph theory. Moreover, this proposed method controls the complexity of algorithm effectively by changing the clustering object from the pixels to gray levels.

The experimental result looks good even on some abnormal case. However, we should have post processing after segmentation to remove some noise and normalize the text object. It's possible to using rgb color level instead of gray-scale level for this method.

REFERENCES

[1] R. Wu, J. Huang, X. Tang, and J. Liu, "A Text Image Segmentation Method Based on Spectral Clustering", *Computer and Information Science, 2008*

[2] Jianbo Shi and Jitendra Malik.(2000) "Normalized Cuts and Image Segmentation." IEEE Transactions on Pattern Analysis and Machine Intelligence

[3] Lienhart R and Wernicke A.(2002) "Localizing and Segmenting Text in Images and Videos." *IEEE Transaction on Circuits and Systems for Video Technology*