Image Region Segmentation Using Normalized Cut

Pengbo Li
Group Oregae
u6683977@anu.edu.au

Quning Zhu
Group Oregae
u6693699@anu.edu.au

Weifan Jiang
Group Oregae
u6683698@anu.edu.au

Abstract

Image segmentation has become a hot research content in the fields of image understanding and analysis and computer vision because of its high application value in many real-life scenarios at present. Traditional image segmentation methods mainly include threshold method [11], boundary detection method [8], region method [1], etc. The implementation principles of these methods are different, but the basic. Originally, it uses the low-level semantics of the image, including the color, texture and shape of the image pixels [15], and the actual segmentation effect is endless when encountering complex scenes. In the paper, we explore the combination of the K-means clustering algorithm and the Normalized Cut algorithm on segmenting images. The features used in this algorithm are color and the locations of the pixels. The location of pixels is benefit to define different regions in images. The result is slightly inferior to ground truth, which is semantics segmentation, but more closed to the original image.

1 Introduction

In recent years, machine learning technology on segmentation tasks has attracted people's attention. Semantic Segmentation on image is to label each pixel in the target category on the image according to "semantics", so that different kinds of things can be distinguished on the image, which allow machine to understand image in pixel level. In addition, Semantic Segmentation widely used on industrial files. For example, self-driving cars have gradually become possible. Autonomous driving is a complex robotic task that requires perception, planning and execution in a constantly changing environment. Since its safety is of utmost importance, it is also necessary to perform this task with the highest accuracy. Semantic segmentation provides information about free space on the road, as well as detecting information such as lane markings and traffic signs.

In the entire machine learning process, algorithms are required to recognize and learn the images provided as raw data. In this process, semantic segmentation technology is applied.

In the early days, the initial application requirement of computer vision was only to identify basic elements, such as edges (lines and curves) or gradients. However, only through the creation of full-pixel semantic segmentation to understand pixel-level images, it brings together the image parts belonging to the same target, thereby expanding the application scenarios of semantic segmentation.

Divide the image segmentation algorithm into bases The three types of graph theory methods, pixel clustering methods and semantic segmentation methods are introduced separately.

Preprint. Under review.

1.1 Review of standard algorithms

The method based on graph theory uses the theories and methods in the field of graph theory, and regards the problem of image segmentation as the problem of dividing the vertices of the graph. The general method is to The image to be segmented is mapped to a weighted undirected graph. Each pixel in the original image is mapped to a vertex of the undirected graph, and the characteristic information of the pixel itself corresponds to the vertex attribute of the graph; the pixel The adjacent relationship between is mapped to the edge of the graph, and the similarity or difference between pixel features corresponds to the weight of the edge. By constructing and solving the cost function of graph information, the vertex set of the graph is divided into foreground set F and background set B.

The spectral clustering method is based on the theory of spectral graphs. By constructing the Laplacian matrix of the original graph and solving the eigenvalues and eigenvectors, the vertices in the graph are separated from the front and the background to solve the image segmentation problem. Wu et al. [16] converted the optimal segmentation problem of an image into solving the minimum cut problem of the corresponding image according to the principle of the equivalence of the maximum flow and the minimum cut in the network. And define cost function as:

$$cut(A,B) = \sum_{u \in A, v \in B} w(u,v) \tag{1}$$

The minimum cut method considers maximizing the difference between subgraphs, but ignores the principle that the internal difference of subgraphs should be as small as possible, and tends to separate individual nodes in the graph. Shi et al. [14] proposed a Normalized Cut (NCut) algorithm in 2000, and normalized cost function1, taking into account the differences between the sub-graphs, while considering the similarity within the sub-graphs and define as:

$$Ncut(A,B) = \frac{cut(A,B)}{assoc(A,V)} + \frac{cut(B,A)}{assoc(B,V)} \tag{2}$$

The NCut algorithm minimizes the cost function globally through the contour features and texture features of the image, and can generate a more regular image block. Shi et al. converted the optimization cost function is into solving the eigenvalue problem of the Laplacian matrix, which belongs to NP complete problem. For this, Ren et al. [12] proposed to use the NCut algorithm to group the image into several larger sub-images, and use the K-means algorithm for each sub-image to further refine the initial division, thereby reducing the time complexity of the algorithm. Xu et al. [7] First, condense the vertices, use the condensed graph sequence to approximate the original image, and then use the NCut method to generate coarse-scale segmentation. Finally, according to the statistical characteristics of the image, use the maximum posterior probability of the mixed model to optimize the segmentation results to obtain the fine-scale segmentation results. Li et al. [9] proposed a linear spectral clustering (LSC) algorithm, which is based on the cost function of the K-way NCut algorithm [14]. Mapping to a high-dimensional feature space, by proving that the cost function of the weighted K-means algorithm and the K-way NCut algorithm share the same best advantage, iteratively use the K-means algorithm to cluster in the high-dimensional feature space Instead of solving eigenvalues and eigenvectors in NCut algorithm.

For the use of Markov random fields to model the image at the pixel level, with the graph cut theory algorithm, using the equivalent relationship of Gibbs distributions, through the baysian maximum a posterior estimation solve the cost function, Greig et al.[6] proposed graph cut that the min-cut/max-flow algorithm can be used to optimize the cost function. In addition, Pixel-based clustering method initialize a rough cluster, and then use an iterative method to cluster pixels with similar spatial distances, color, brightness, texture and other features to the same superpixel, until convergence, so as to obtain the final image segmentation result. Even most popular end-to-end deep learning method, they are all making outstanding contributions to the in the research of image segmentation.

2 Related Work

Discussions about segmentation and clustering can be dated back over 40 years. There have been various explorations of the algorithm applying for image segmentation. The application of these algorithms shows up in different areas. In this section, we briefly cover some of the work.

2.1 Region splitting and merging method

This is an early approach to image segmentation. This algorithm iteratively splits and merges the regions to form the optimal segmentation. Merging considers the similarity, while splitting considers the heterogeneous. The splitting and merging follow some uniformity criteria which obey a subset property [3]. When the uniformity predicate is true for some sets, it should also be true for the subset of these sets [5].

2.2 Mapping image pixels into some feature spaces

The method proposed by Comanciu and Meer [2] firstly transfers the points in the same cluster closer together. Smoothing operations helps the transformation. Then the algorithm expands each point to a hypersphere. The clusters are the connected components of the expansions. This method also works when the points of the cluster do not lie within a fixed distance.

2.3 Methods relating to the Normalized Cut

Wu and Leahy [16] proposed an algorithm by finding the minimum cuts in a graph. Minimizing the similarity between the boundary pixels is the cut criterion. However, this method is limited to find small components. The Normalized Cut algorithm can address this bias. After the Normalized Cut algorithm, much research combined it with other algorithms to solve the inefficiency problem. The design of Makrogiannis et al. [10] firstly applied the watershed algorithm on images to get a graph structure. Then the authors used the Normalized Cut algorithm to partitioned the graph structure.

3 Method description

This section has three sub-sections, mainly introduces methods we used in image region segmentation based on normalized-cut.

3.1 K-means

We use K-means to cluster the pixels in the image into several classes, and each class represents a node of the region adjacency graph constructed in the next step. For better effect, we use the K-means++ method to find the initial clustering centroids for K-means.

3.1.1 K-means++

To reduce the potential risk of randomly selecting the initial cluster centroids, we use the K-means++ method to select the initial cluster centroids with the following steps:

The purpose of this method is to make the initial cluster center as far away as possible, reduce the poor effect or long operation time caused by the terrible location of the randomly selected cluster centroids, improve the final error, and help K-means converge quickly and reduce the calculation time.

3.1.2 The distance between pixels

When calculating the distance of each pixel, as the color and the location of the pixel itself will both affect the classification performance, we consider the color and location comprehensively. Moreover, considering that the influence of color and location is not necessarily the same, we also attach different weights to color and location.

First of all, we convert the three-dimensional image we read in with m * n * c shape into a two-dimensional matrix with shape N*c, which is called X, m,n c is the width, length, and the number of channels for the image. N is the total number of pixels in the image and it is easy to see that

$$N = m * n$$

In matrix X, each row represents the RGB channel value of one pixel, and each column represents the corresponding value of all pixels in one channel. Considering that the location of pixels also influences region segmentation, we add the index of row and column of each pixel to matrix X. In addition, as the influence of color and location may not be different, we decided to give different weights to the columns representing color and position. The specific weights are described in the next section.

We use the L2 norm to calculate the distance between two data points:

$$distance = \|\mathbf{x}-\mathbf{y}\| = \sqrt{(r_x - r_y)^2 + (g_x - g_y)^2 + (b_x - b_y)^2 + (row_x - row_y)^2 + (col_x - col_y)^2}$$

,where x, y represent any two processed pixels respectively, r_i, g_i, b_i represent the corresponding value of pixel I on the RGB channel, Row_i, col_i represent the index of the row and the column of pixel i.

3.1.3 K-means

We use the initial centroids selected in subsection 1 and the distance calculation method mentioned in subsection 2 to do K-means clustering for the processed data, the purpose is to preliminarily divide the image into some regions formed by similar pixels to obtain the region adjacency graph of the image.

K-means is a widely used clustering algorithm, when dealing with large data sets, this algorithm can ensure good scalability. Because its performance is affected by the selection of the initial centroid, we use K-means + + method to get the initial centroid, which makes the distance of the initial clustering centroid as far as possible to reduce the negative impact caused by the bad location of the initial centroid.

The main purpose of this method is to prepare for the next step of constructing the region adjacency graph.

3.2 region adjacency graph

We construct a region adjacency graph (RAG) based on the result of K-means clustering to do further normalize-cut.

Region adjacency graph is a typical topological data structure. Topological data structure describes an image as a group of elements and their relationships, which are usually represented by a graph structure. A node in the graph corresponds to a region in the image. Each node contains a set of similar pixels. When there is a common boundary between the two regions, it means those 2 regions have an adjacent relationship and the relationship will represent by an arc. We divide the region according to the results of K-means in the previous section and then obtain the nodes and arcs of RAG.

The region adjacency graph explicitly stores the adjacent information of all regions in the graph. The image is divided into a set of regions composed of similar pixels, and these regions are represented by a graph structure. The new graph node no longer represents a single pixel, but a pixel region. In this way, the size of the graph is reduced while the features of the graph are retained as much as possible, which greatly reduces the computing time and space of NCUT in the next step.

3.3 Global recursive normalized-cut for multiple classes

Finally, we use the recursive normalized-cut method to divide the RAG into eight classes, and then get the final label of each pixel according to the class label of each region.

Sarkar et al. [13] improved the NCut algorithm and proposed the average cut (ACut) algorithm and define the cost function as :

$$Avcut(A, B) = \frac{cut(A, B)}{|A|} + \frac{cut(B, A)}{|B|} \tag{3}$$

,where |A| and |B| represent the number of vertices in subsets A and B, respectively. This cost function represents the sum of the ratio of the boundary loss to the area of the region in the graph, and minimizing the cost function can produce accurate Division. The shortcomings of the ACut algorithm

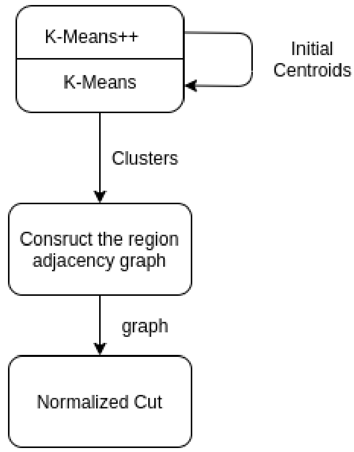

Figure 1: Construction of our work flow.

are similar to the NCut algorithm, and it is easy to segment a smaller subgraph containing only a few vertices. Ding et al. [4] proposed a min-max cut (MCut) algorithm to solve the shortcomings, and defined the cost function as:

$$Mcut(A, B) = \frac{cut(A, B)}{W(A)} + \frac{cut(B, A)}{W(B)} \tag{4}$$

where W(A) is defined as the sum of the weights of all edges in subset A. According to the principles of minimum inter-class similarity and maximum intra-class similarity, it is necessary to minimize cut(A,B) and maximum Reduce W(A) and W(B), the minimization formula (4) can segment a more balanced cut set.

4 Experiments

In this section, we evaluate the performance of the combination of the K-means and the Normalized Cut algorithm on selected ten images from the IDD-Lite dataset. Firstly, we describe the experimental protocol and the dataset. The construction of the K-means and the Normalized Cut follows. Based on the algorithms, we segment the image into eight classes and compare our results with the given labeled images. Finally, we discuss the merits and demerits of our design.

4.1 Dataset and experiment setup

The dataset used in this experiment is the IDD-Lite dataset. Images for segmentation in this dataset contain the streets, passersby, and various transportations. The dataset also provides the labeled images. The labels are assigned based on the semantics. Semantic segmentation means the same objects belong to one group and have the same color in the labeled images. We select ten images from the dataset. They are shown in Fig. 3. The upper line in Fig. 3 shows the ground truth images. Our results are in the bottom line in Fig. 3.

The algorithm used on images contains two parts. The first step is implementing the K-means algorithm. To avoid the bias of the random initialization in the K-means algorithm, we adopt K-means++ to decide the initial centroids. To simplify the processing, we set the number of clusters the K-means should separate as 16. These clusters became the nodes in a region adjacency graph. We examine the locations of each two nodes. If nodes are adjacent to each other, we add one edge into the graph. Finally, we use the Normalized Cut algorithm on the region adjacency graph to get the final segmentation. The construction of our work flow is shown in Fig. 1.

4.2 Design explorations

Firstly, we attempted to use the Normalized Cut on the original images. The size of the given image is 227 times 320. The number of pixels reaches more than 70000. Although our computers can create

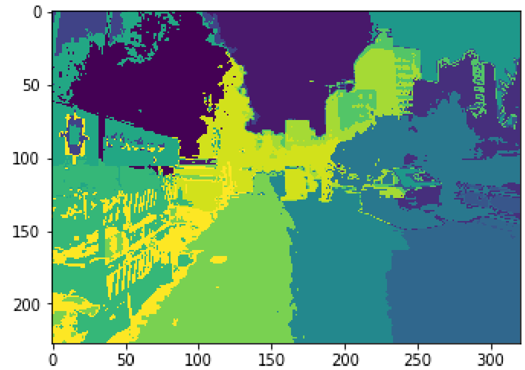

Figure 2: The result generated by the K-means

the matrix in this size, they cannot handle the matrix calculations. Also, the eigendecomposition on this large matrix is time-consuming. Under this situation, we tried to adopt a simple clustering algorithm to do the down-scale before using the Normalized Cut algorithm. K-means is our choice for its convenience and efficiency.

The features we initially choose are the color of each pixel. The Euclidean distance is the measurement of the similarity of two pixels. However, only considering the color results in the scattered clusters. The result after K-means is mottled–some pixels are far away from each other but have the same label. To get the region segmentation, we add the location of pixels as additional features for K-means. Then we get more gathered results.

After K-means, we obtain sixteen clusters. The plot of the clusters is like Fig. 2. Each of the clusters is the node in the region adjacency graph. The mean color value of all pixels in one cluster is the feature of the node. As we can examine the locations of pixels in different clusters, we can know whether two clusters are neighbors. When constructing the graph, we add edges between two adjacent nodes rather than all nodes. The weights are the similarity of the two ends based on the mean color.

4.3 Discussion on results

Fig. 3 shows the selected ten images and the corresponding labeled images and results. Images in the first and the fourth line are the original images; the second and the fifth line shows the provided labeled images, while the second and the fourth line show the results generated by our algorithm. As mentioned above, the labeled images consider the semantic of the area. We can find people are in one color and the buildings are in another color. The Normalized Cut algorithm generates results that are more closed to the original images.

The results of these two clustering methods may have different purposes based on application situations. When we consider object detection, the labeled image may be suitable than our results. But for image blurring or down-scaling, the results in this experiment may have more advantages.

4.4 Discussion on algorithm design

The algorithm proposed here, taking the locations and the color as features, generates the images that are closed to the original images. The K-means algorithm efficiently downscales the image. The Normalized Cut algorithm works on the region adjacency graph to generate final clusters. The combination of these two algorithms performs stable on different kinds of inputs. However, the connection between these two parts needs more considerations. Sometimes the results from the K-means algorithm cannot construct a reasonable region adjacency graph for the Normalized Cut. The illogical clustering from the K-means may result in the merging of several parts. Then we cannot obtain eight classes from the Normalized Cut. As we cannot avoid the randomness of the K-means, we need to be more careful with this part.

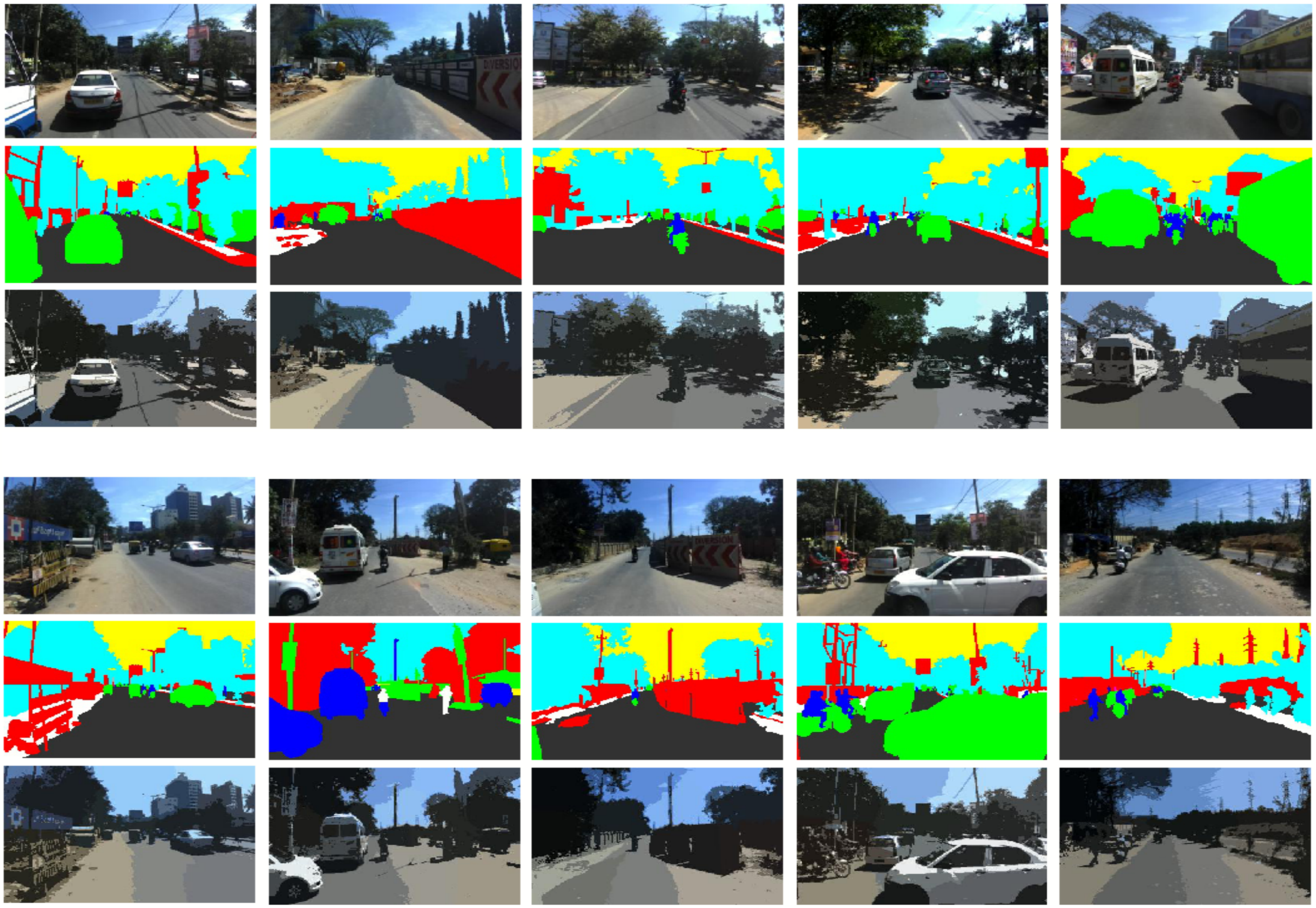

Figure 3: Comparison between the labeled images and our results. The images in the first and fourth line are the original images. The images in the second line and the fifth line are the labeled images. The images in the third and sixth line are our results.

5 Conclusion

In this report, we implement the normalize-cut algorithm on the IDD-Lite dataset. The IDD-Lite dataset contains RGB images and labeled images. We use the RBG images to do normalized-cut. In the process of normalize-cut, we first apply K-means clustering to the image to construct the region adjacency graph (RAG) to represent the image. Then we use normalize-cut on RAG instead of the original image. The label of each nodes is the result from K-means. The value of each node is the mean color value of the clusters defined after K-means. We consider the locations of each clusters and use the mean color of each node to generate edges of the RAG. This method greatly reduces the computing time and memory and has better performance than using normalize cut directly on the original image.

Through the comparison between our results and label, we can see that we have successfully divided the similar colors in the image into the same area, but there is still a certain gap compared with the label. The labeled images segment areas based on semantics. It is difficult to achieve the same performance as the label in the dataset only by applying the Normalized Cut based on colors, as the same object may have multiple colors.

In the future, we can continue to improve the algorithm and take more ways to segment the image more accurately. For example, we can use the SLIC algorithm instead of K-means when constructing RAG. SLIC algorithm can generate compact and nearly uniform superpixels and has high comprehensive evaluation in terms of operation speed, object contour preservation, and superpixel shape. Therefore, it is a very worthwhile algorithm to consider.

In addition, considering the characteristics of the label in the dataset, for dividing the items more accurately in the image, we can also adopt semantic segmentation techniques such as train the neural network to segment the image more accurately.

References

[1] R. Adams and L. Bischof. Seeded region growing. *IEEE Transactions on pattern analysis and machine intelligence*, 16(6):641–647, 1994.

[2] D. Comaniciu and P. Meer. Mean shift analysis and applications. In *Proceedings of the Seventh IEEE International Conference on Computer Vision*, volume 2, pages 1197–1203. IEEE, 1999.

[3] M. C. Cooper. The tractability of segmentation and scene analysis. *International Journal of Computer Vision*, 30(1):27–42, 1998.

[4] C. H. Ding, X. He, H. Zha, M. Gu, and H. D. Simon. A min-max cut algorithm for graph partitioning and data clustering. In *Proceedings 2001 IEEE international conference on data mining*, pages 107–114. IEEE, 2001.

[5] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient graph-based image segmentation. *International journal of computer vision*, 59(2):167–181, 2004.

[6] D. M. Greig, B. T. Porteous, and A. H. Seheult. Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society: Series B (Methodological)*, 51(2):271–279, 1989.

[7] X. Hai-Xia, T. Zheng, and D. Ming-Tao. Multiscale segmentation for sar image based on spectral clustering and mixture model. *Journal of Image and Graphics*, 15(3):450–454, 2010.

[8] S. Lakshmi, D. V. Sankaranarayanan, et al. A study of edge detection techniques for segmentation computing approaches. *IJCA Special Issue on "Computer Aided Soft Computing Techniques for Imaging and Biomedical Applications" CASCT*, pages 35–40, 2010.

[9] Z. Li and J. Chen. Superpixel segmentation using linear spectral clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1356–1363, 2015.

[10] S. Makrogiannis, G. Economou, and S. Fotopoulos. A region dissimilarity relation that combines feature-space and spatial information for color image segmentation. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 35(1):44–53, 2005.

[11] N. Otsu. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1):62–66, 1979.

[12] X. Ren and J. Malik. Learning a classification model for segmentation. In *Computer Vision, IEEE International Conference on*, volume 2, pages 10–10. IEEE Computer Society, 2003.

[13] S. Sarkar and P. Soundararajan. Supervised learning of large perceptual organization: Graph spectral partitioning and learning automata. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(5):504–525, 2000.

[14] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence*, 22(8):888–905, 2000.

[15] A. W. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Transactions on pattern analysis and machine intelligence*, 22(12):1349–1380, 2000.

[16] Z. Wu and R. Leahy. An optimal graph theoretic approach to data clustering: Theory and its application to image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 15(11):1101–1113, 1993.