Lab 7: BSDS500 Benchmarks

Cristian M. Amaya
Universidad de los Andes
cm.amaya10@uniandes.edu.co

Abstract: In this laboratory, a segmentation function previously created will be evaluated in the BSDS500 dataset with its own evaluation methods, to test the function some parameters as segmentation method, color spaces and number of clusters will be chosen to obtain segmentation results to compare against the human annotations.

Keywords: Segmentation, BSDS500, k-means, GMM

1.. Introduction

The Berkley Segmentation Dataset (BSDS) has been an important benchmark in the investigation history of artificial vision, as it was free to the public and any investigator in the world could use this images and human annotations to compare the results from their methods with peers from different institutions. The first dataset consist of 300 images with a different number of human annotations each, the one that will be used to test the segmentation function will be the BSDS500 that contains 500 images with corresponding anotations.

In this laboratory, the previous segmentation function developed will be evaluated using the evaluation methods develop for the BSDS500, this in hopes to obtain more accurate results that the simple segmentation evaluation used in the previous lab.

2.. Methods and Materials

2.1.. Segmentation function

This function was developed in the previous lab, it is able to use four different segmentation methods such as k-means, GMM, watershed transform and the hierarchical segmentation method in the three standard color spaces: RGB, HSV and L*a*b*. Additionally, the segmentation method can consider spatial information, integrating the pixel coordinates as channels in the input image and a number of clusters must be specified to obtain at least that number of segmentations in the output label matrix.No changes were applied to the function, the results were compatible with the database.

2.2.. Segmentations and Parameters

To test the segmentation function, the 500 images in the BSDS500 dataset were segmented using the two segmentation methods, with their respecting best color spaces, that produces the best Jaccard scores. In this case, k-means with the HSV color space and GMM with RGB color space and XY coordinates were used, as their produces the best results in the previous lab, and a set of numbers of clusters was also specify.

First, a set of 3 numbers of clusters (5,7 and 9) was used to produce a total of 6 segmentations per image in the dataset and were saved in the specific format of the benchmarks functions, then a new set of numbers of clusters were used (5 to 10) to produce a total of 12 segmentations using both methods, this two sets of segmentations were used to evaluated the function.

2.3.. Evaluation

Two different evaluation methods for the BSDS500, a fast evaluation and a more complete evaluation, were used to evaluated the regions segmentation and boundary detection. Both methods were using alongside the two sets of segmentations to produce the results of the segmentation function in this dataset.

3.. Results and Discussions

3.1.. Fast Evaluation Method

As one can observe in figure 1, the boundary results are very low, this can occur as the fast boundary detection algorithm uses morphological operations to produce a more accurate result that a straight up comparison, but this method can produce a large number of false negatives as the boundary are a small portion of the total image. The region segmentation results were expected, as using a predefine set of clusters can produce a low number of segmentations in comparison to the human annotations, it is still a good result for the most basic and reliable methods of segmentations such as k-means and GMM.

As its show in the figure 2 and 3, that are the Precision-Recall curves of the test images using 6 and 12 number of clusters respectively.

3.2.. Complete Evaluation Method

Using this more complete evaluation method produces better results in the boundary detection evaluation, as it can be seen in figure 4, this could had happened as this evaluation attempts to make the boundaries more similar to produce a better result, as boundaries can vary depending on the human annotation. The region segmentation results are the same as in the previous method, but this results are similar to the ones produces in the boundary detection which is a good indicator.

The segmentation method is more reliable than expected, using this can produces good segmentations, not as good as the probability of boundary (pB) or the global pB method that take more information into account, but are better that expected using k-means and GMM as the central methods.

In figure 5 and 6, the precision-Recall for the test images using 6 and 12 number of clusters respectively, in comparison to the previous curves, one can perceptually conclude that the segmentations methods can produce a better detection result that the ones observe in the fast evaluation method. Ideally, the Precision-Recall curve should be a curve with F-max near 0.79 and being consistent along high precision and recall values, but with simple methods this could be impossible to achieve.

3.3.. Best Method

In the figure 7, the results of the complete evaluation method for k-means and GMM segmentations, as it can be observe GMM with the RGB color space and XY coordinates produces better results both in boundary detection and in region segmentation. This could be possible as the method takes into account spatial information which can facilitate the clustering of nearby objects.

In figure 8 and 9, the Precision-Recall curves for both test sets are shown, K-means provides greater recall values, but GMM produce higher precision values while losing little recall. This could happen as GMM with XY coordinates can produces results with lesser false negatives than k-means without spatial information.

4.. Conclusions

This evaluation in the BSDS500 database is a good approximation to the labor that many investigators made years ago, to be able to achieve a better score than others, it's a good method of comparison between peers and algorithms, one can compare the segmentation methods to observe which can produce the better result in the least amount of time and optimize the methods and parameters.

The segmentation method created in the previous lab produce better region segmentations results using this evaluation methods in comparison to the simple Jaccard Index comparison used before, and it can be observe as a good method for simple segmentations as it can produces a good approximation to human segmentations taking into account its simple methods and the information they use.

In the other hand, the function and its methods are not the best approximation as they are much better methods that not only take into account spatial and intensity information, like pB or gpB,that additionally do not require to define a specific number of clusters, which is a crucial parameter in the basic methods.

Images

Fast Evaluation Method

# of segmentations	Set	Boundary		Region		
		ODS	OIS	ODS	OIS	Best
6	Test	0,2	0,21	0,42	0,47	0,52
	Train	0,20	0,22	0,43	0,48	0,53
	Val	0,19	0,2	0,42	0,46	0,51
12	Test	0,20	0,22	0,42	0,49	0,55
	Train	0,20	0,22	0,43	0,50	0,56
	Val	0,19	0,20	0,43	0,48	0,54
Mean		0,20	0,21	0,43	0,48	0,54

Figure 1. Results for the Fast Evaluation Method

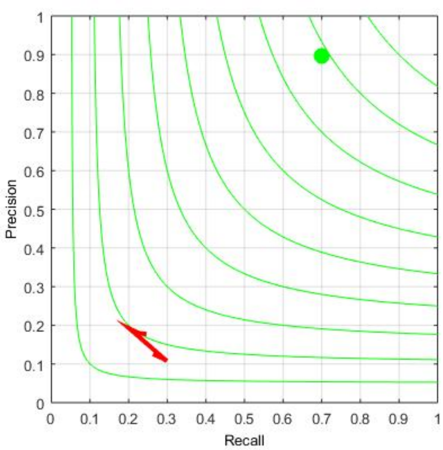

Figure 2. Precision-Recall Curve for 6 segmentations

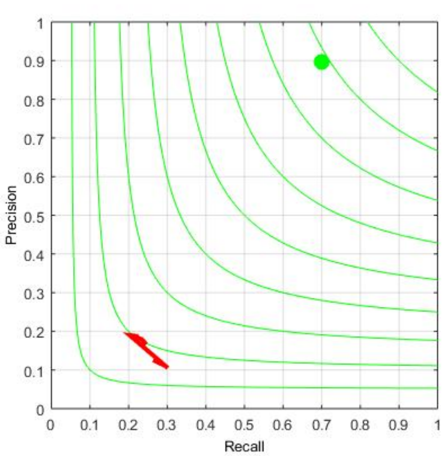

Figure 3. Precision-Recall Curve for 12 segmentations

Complete Evaluation Method

# of segmentations	Set	Boundary		Region		
		ODS	OIS	ODS	OIS	Best
6	Test	0,52	0,55	0,42	0,47	0,52
	Train	0,51	0,55	0,43	0,48	0,53
	Val	0,51	0,54	0,42	0,46	0,51
12	Test	0,52	0,57	0,42	0,49	0,55
	Train	0,51	0,56	0,43	0,50	0,56
	Val	0,51	0,55	0,43	0,48	0,54
Mean		0,51	0,55	0,43	0,48	0,54

Figure 4. Results for the Complete Evaluation Method

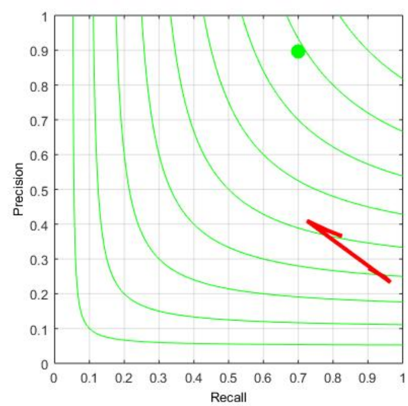

Figure 5. Precision-Recall Curve for 6 segmentations

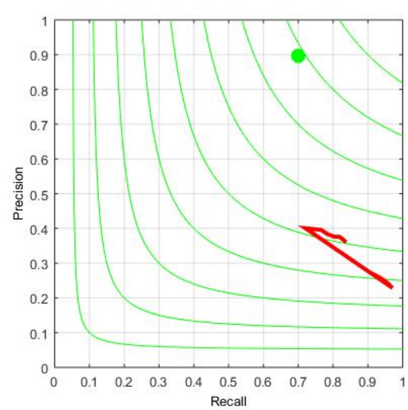

Figure 6. Precision-Recall Curve for 12 segmentations

K-means vs. GMM

Method	Set	Boundary		Region		
		ODS	OIS	ODS	OIS	Best
K-means	Test	0,42	0,42	0,38	0,39	0,41
	Train	0,40	0,40	0,39	0,40	0,41
	Val	0,39	0,40	0,36	0,37	0,38
Mean		0,40	0,41	0,38	0,39	0,40
GMM	Test	0,52	0,55	0,43	0,46	0,50
	Train	0,51	0,54	0,43	0,47	0,50
	Val	0,50	0,53	0,42	0,45	0,49
Mean		0,51	0,54	0,43	0,46	0,50

Figure 7. Results with the complete evaluation method compare K-means and GMM

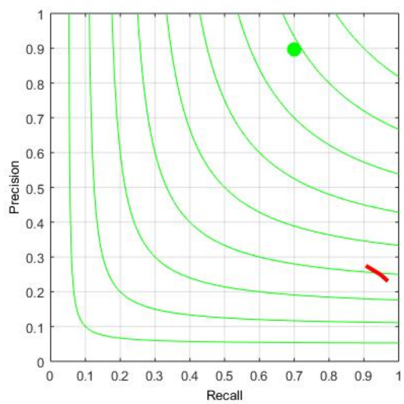

Figure 8. Precision-Recall Curve for 3 segmentations with K-means

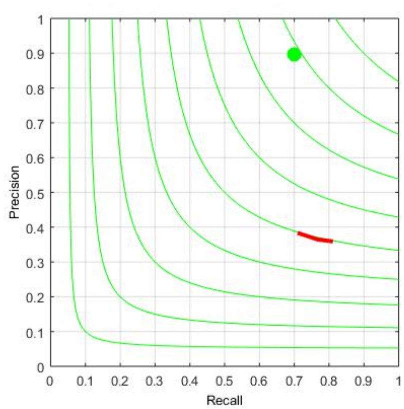

Figure 9. Precision-Recall Curve for 3 segmentations with GMM