

Laboratory 5: Berkeley Segmentation Dataset and Benchmarks 500 (BSDS500)

Javier Coronel
Universidad de los Andes
Biomedical Engineering

jd.coronel30@uniandes.edu.co

Luis Carlos Rivera
Universidad de los Andes
Biomedical Engineering

lc.rivera10@uniandes.edu.co

Abstract

Segmentation is one of the main areas of computer vision, the principal objective is to obtain from an original image, each object represented as one area. Nowadays there are different approaches to solve this problem and each method have pros and cons, but there is no perfect approach for the problem. In this case, an implementation of a segmentation function was done using different methods and color space representations. The methods used for segmentation were k-means, watersheds, hierarchical and GMM. The function was used in the BSDS500 and to select the best algorithms and color space representation, different configurations of the function were developed in the training dataset. After a qualitative approach, lab color space, k-means and hierarchical segmentation were selected to run over the test dataset. The k-means algorithm produces an ODS of 0.52 and hierarchical segmentation an ODS of 0.62, suggesting a best performance for hierarchical segmentation.

1. Introduction

The segmentation of images is a widely studied problem [1, 2], it consist in grouping and partitioning the image in multiple segments. By doing the segmentation, it is expected to obtain a more meaningful image [3]. In the present laboratory there will be analyzed the results of the created function that implements different clustering methods such as K-means, Watershed, Hierarchical, GMM (Gaussian Mixture Models). The function is evaluated on the Berkeley Segmentation Dataset and Benchmarks 500 (BSDS).

2. Segmentation Methods

The implemented function uses four distinct methods: K-means, Watershed, Hierarchical, GMM (Gaussian Mixture Models). It also consider the feature space of representation for the segmentation, these are rgb, lab

and hsv. For the representation, it also has the option of considering spatial information of the input images in their respective color space.

K-means

K-means is a clustering method for data distribution. It models the probability density as a superposition of spherically symmetric distributions. The algorithm receives the data and the number of clusters 'k' to find, then, iteratively updates the centroids (cluster center) based on the surrounding data [4]. Finally, until a convergence is reached, the resulting centroids generate the data distribution.

Watersheds

One way to interpret the watershed algorithm is to start flooding the image from the local minima (assuming the image like a landscape) and labeling where different regions meet[4]. The general algorithm uses the gradient magnitude of a gray-level image, and based in the local minima start the segmentation. In this case, watershed algorithm was implemented using *h-minima*. This way, the imposed minima in the gradient were not local but hierarchical.

GMM

The Gaussian Mixture Model is a soft clustering method, this means that, for each point, it assigns a probability of belonging to cluster. The goal of the GMM is to identify the correct parameters to explain the data as Gaussian distributions. Like *k-means* algorithm, it needs to specify the number of clusters.

Hierarchical

The hierarchical segmentation uses perceptual features, the more general used is homogeneity in color tone. At the initial state, the algorithm start with an over segmentation where every single pixel represent a region but this clustering method have the characteristic of leveling up, each level is defined as a distance where now, the new

regions are redefined by the perceptual parameter. As the algorithms scales up it reaches a critical point where the segmented areas are well defined, but any value over it is over-segmentation and under is over-segmentation. The disadvantage of this method it's the computational resources that it takes for big images and big datasets like BSDS500 [5].

Selected method

To select the best algorithm, each model was implemented in the training dataset using different configurations for the parameters. Then, based in image results (Fig. 1), the best configuration of parameters and methods were selected. The final selected algorithms were k-means and hierarchical segmentation with the lab color representation. This two methods presented the best behavior over the training set (computational time, memory requirements, visual results).

As mentioned above, there were several specifications to select the color space for the segmentation. The selected color space representation was lab. This, because in the training set, this color space was the one that represented different regions in each image in the best way.

3. Methodology

To test the general performance of the function, we used the BSDS500 dataset to determinate which of the cluster methods worked best or worst. The reason of using BSDS500 Dataset is because:

- It uses natural images.
- It implemented human annotations over each image.
- It allowed a quantitative evaluation over the final segmentation work.

The dataset is composed by 500 images, separated in three different categories: training, test and validation, each one of this with 200, 200 and 100 images respectively. Is important to clarify that BSDS500 uses BSDS300 as training. All the images where in RGB but with different size. The total weight of the dataset including benchmark codes is 251 Mb. As mentioned before one of the main reasons for using this dataset was the annotations made by humans, this annotations are represented in the folder of groundtruth. Each of the files in the folder represent one image with the same name, the difference is that in the groundtruth each pixel has been marked with a number to represent the region of origin.

The main idea of the benchmarks is to make a binary classification (boundary/not boundary pixel). Also it evaluates the quality of the segments, the previous criteria have

been evaluating using 3 different approaches, the first one denominated *Variation of Information* which evaluates the distance from a segment obtained from the clustering Vs the groundtruth region. The second *rand Index* which gives the probability of getting a random index pixel in the segmentation compared with the groundtruth. Finally the *Segmentation Covering* which gives an approximate measure of how well the machine segmentation can explain the human segmentation. The implemented function have the following structure:

```
my_segmentation = segment_by_clustering(rgb_image,
feature_space, clustering_method, number_of_clusters)
```

Where the inputs are an image in RGB, a feature space if the user wants to change from RGB to any other of the following ones (Lab, HSV, RGB+xy, HSV+xy, Lab+xy), clustering method the available ones are (K-means, Hierarchical, GMM, Watershed) defined by the user and finally an optional parameter the number of clusters for methods that require an specific input number.

its important to remark that for the definition of which methods were the best and the parameters we used both of our functions due to the differences of the implementation in some methods.

4. Results

Over the training dataset, the function produced different results for the different algorithms using lab colorspace. The results are reported in figure 1. There, it can be seen that, despite over segmentation in various algorithms (e.g. watersheds), the lab color space is useful for segmentation in k-means and hierarchical segmentation. One of the most accurate results of k-means and hierarchical segmentation in the test BSDS500 are reported in figure 2. There is evident that there is a difference between groundtruth and segmentation, but in general the main regions are well divided. Is noticeable, that the contours of the segmentation presents distortion and noise. This result in a displacement of the boundary and may have significance in the quantitative results of each algorithm.

In figures 3 and 4 are the Precision-Recall curves. It can be seen that, for both methods the behavior is similar, following an almost linear decay. For the K-means PR curve, at the begining is a low value, indicating a big number of false detection. But then, its behavior normalizes. The PR curve for hierarchical segmentation reaches a better performance with a more linear decay.

5. Discussion

First of all, one color space is not better than any other, it only depends on the purpose of the algorithm. In this case, the lab color space representation gave better results than

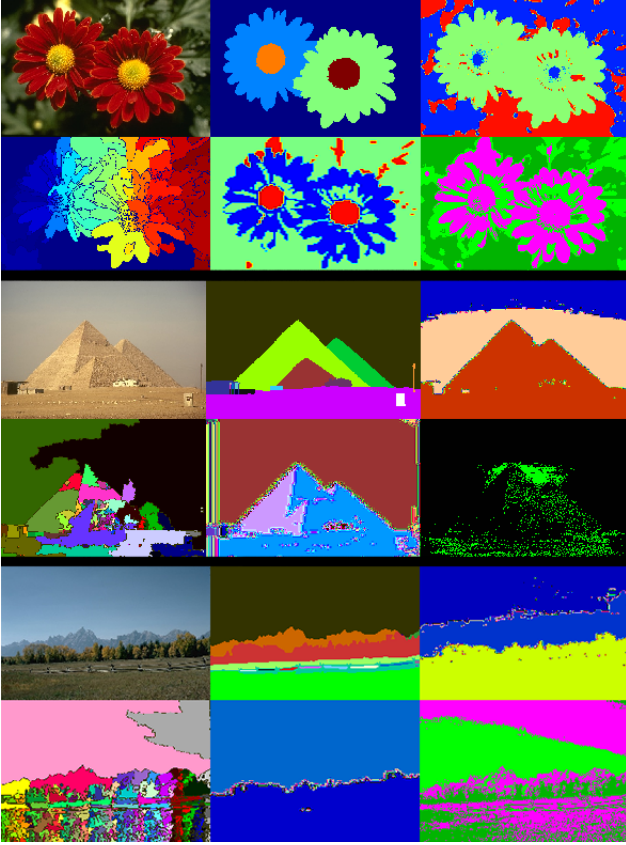


Figure 1. Results over three different images on training dataset. Top left: Original, Top center: Groundtruth, Top right: K-means, Bottom left: Watersheds, Bottom center: Hierarchical, Bottom right: GMM



Figure 2. Example of results over test dataset after selecting the algorithms. First: Original, Second: Groundtruth, Third: K-means, Fourth: Hierarchical

the others. As specific advantage, this color space allow the algorithm set a real difference between color tones in each region.

As relevant note, we determine that GMM was a good method for clustering; but it was a disadvantage related with higher number of clusters. It requested excessive memory and more computational power as the number of clusters in-

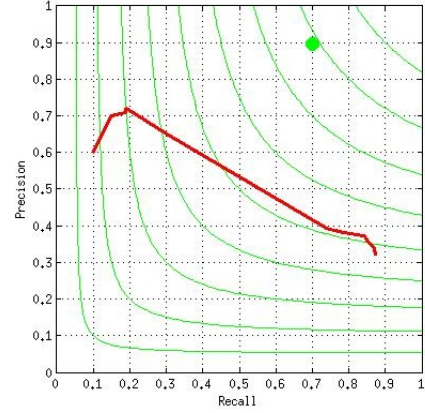


Figure 3. This plot presents Precision-Recall for K-means algorithm

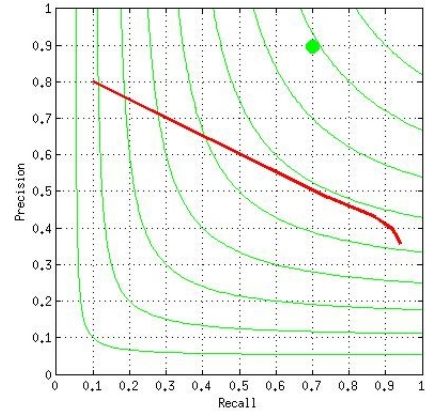


Figure 4. This plot presents Precision-Recall for Hierarchical algorithm

creases. So, in a great variety of images it didn't work and the method was declassified.

Comparing the Optimal Dataset Scale (ODS) and the of the algorithms (table 1), we can infer that hierarchical segmentation has a better performance. The difference between the selected algorithms lies in the initialization.

The limitations of using k-means is that is necessary know the number of clusters. With the k clusters, the first centroids are generated randomly. This may result in different segmentations in different runs of the algorithm. Because of this, k-means does not give a general solution for all the images, in some images works really well with a great recall identifying all the segments but no with a great precision. This is specially clear in the cases in which the number of clusters in the image is bigger than the segmented regions in the groundtruth. Also is important that when the generated clusters in the featured space representation is not spherical, k-means has no good performance according to its euclidean distance.

For hierarchical segmentation, this method has no ran-

dom initialization, it produces more consistent results and is more repeatability than k-means over all the dataset. Because of this, is recommended to use Hierarchical segmentation.

Method	ODS	OIS
Hierarchical	0.62	0.64
K-means	0.52	0.54

Table 1. Segmentation results over BSDS500.

6. Future Work

To improve the result of the segmentation, is recommended a more exhaustive work in the selection of the algorithm. Not only selecting a limited number of images, but also because identifying categories in which the algorithm gives better results or worst. For example identifying geometrical figures versus numbers, some methods could work better for one category than for the other. This because there it's not a single method to segment everything. It also could be interesting to identify quantitatively how exactly differ the performance in training for the methods that require an input such as numbers of clusters, distance between roots and methods without any input.

References

- [1] C. Brice and C. Fennema. Scene analysis using regions, 1970.
- [2] D. Cremers, M. Rousson, and R. Deriche. A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape, 2007.
- [3] L. Shapiro and G. Stockman. Computer vision. pages 279–325, 2001.
- [4] R. Szeliski. Computer vision: algorithms and applications. *Springer Science and Business Media*, 2010.
- [5] N. Teixidó, A. Albajes-Eizagirre, D. Bolbo, E. Le Hir, M. Demestre, J. Garrabou, L. Guigues, J. M. Gili, J. Piera, T. Prelot, et al. Hierarchical segmentation-based software for cover classification analyses of seabed images (seascape). 2011.