BSDS

Cristina Isabel González Osorio
ci.gonzalez10@uniandes.edu.co

Raúl Santiago Molina Rodríguez
rs.molina10@uniandes.edu.co

Abstract

In this practice, we selected two segmentation algorithms (GMM and hierarchical clustering) to be tested in the Berkeley Segmentation Dataset (BSD500) [2], generating Precision-Recall curves for each method and comparing them to the UCM method [1].

1. Introduction

To evaluate the performance of any computer vision algorithm, it is necessary to develop appropriate metrics and frameworks to test them. One of these metrics is the Precision-Recall curves (PR curves). If a segmentation or contour detection algorithm defines its output by a threshold, it is possible to evaluate the algorithm in different threshold values, a trade-off between the percentage of true detections in the output (precision) and the percentage of true detections detected overall (recall), more precisely:

$$P = \frac{TP}{TP + FP} \qquad R = \frac{TP}{TP + FN}$$

In our segmentation methods, however, there is not a numerical threshold, for that reason the PR curves were constructed varying the number of image partitions to find (K).

2. Materials and methods

2.1. Clustering algorithms

Mixture of gaussians

A more generalized approach of Lloyd's algorithm, it uses Expectation Maximization (EM). Given a parametrized probability distribution (Gaussian), with an initial guess of said parameters, the partial membership (responsibility) to each distribution (K distributions) of each datapoint is estimated (Expectation), then, given the responsibilities, the parameters are estimated again.

Hierarchical Clustering

The function *clusterdata* of MATLAB implements an agglomerative hierarchical clustering method. In this order of ideas, the clusters with the least distance are merged in a hierarchical way. Starting from the pixels as individual clusters until reaching the number of clusters specified by parameter. In this way, we obtain a family of clusterings represented in a dendrogram which is constructed from individual elements until they are mixed in a single cluster.

These two clustering algorithms were selected due to their performance in previous segmentation tests (Practice 6), although some modifications were required. Due to the great number of images in the BSD, each image was downscaled to a quarter of its size for segmentation and then upscaled back to its original size. Even though this might diminish the quality of the segmentation results (one could argue that it reduces noise too) it was necessary for the methods to run multiple times in the entire dataset.

Another advantage of these methods was its reduced number of hyperparameters, the number of partitions K being one of them. Other possible parameters considered were the representation space (HSV), image size and, in our case, the upscaling method (nearest neighbor was chosen). All of these parameters, except for K, were chosen after testing the methods in a sample of the images in the training set (our validation set).

2.2. Evaluation

2.2.1 Framework

Both methods were run in 20 different values of K, ranging from 2 to 30, each one being stored in a MATLAB cell object and stored in a *.mat* file. To construct the PR curves, the *allBench_fast* function (provided in the dataset) was used: taking as input the directory of the cell objects stored for each image, calculating evaluation values for boundary and regions storing them in *.txt* files in the output directory.

To visualize the resulting PR curves, the *plot_eval* (provided in the dataset) was used. Besides the PR curve for a respective method, the function also plots the PR point associated with human performance; an ideal segmentation method would contain that point in its curve while covering the greatest amount of area possible (average precision).

3. Results

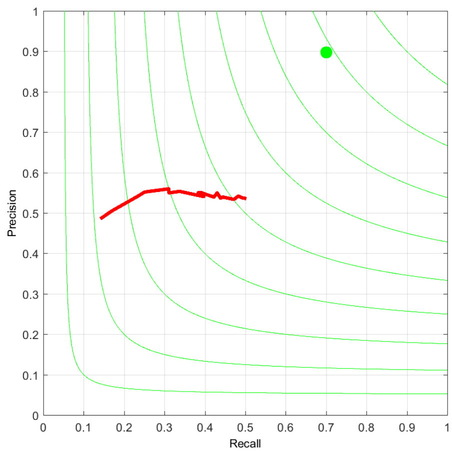

Figure 1: PR Curve for the GMM segmentation method

For the GMM method (Fig. 1), the Optimal Dataset Scale (ODS) for Ground Truth coverage was 0.40 for $K = 15$, the Optimal Image Scale (OIS) was 0.42 and the best result was 0.44.

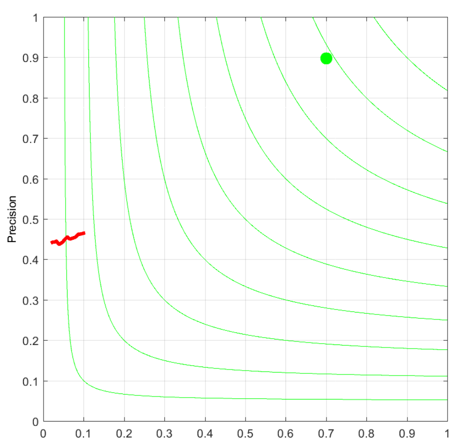

Figure 2: PR Curve for the Hierarchical Clustering segmentation method

For the Hierarchical Clustering method (Fig. 2), the ODS was 0.30, the OIS was 0.31 and the best result was also 0.31.

For comparison, the UCM segmentation method is presented (Fig. 3). It has a ODS of 0.59 (threshold = 0.2), OIS of 0.65 and its best result is 0.74.

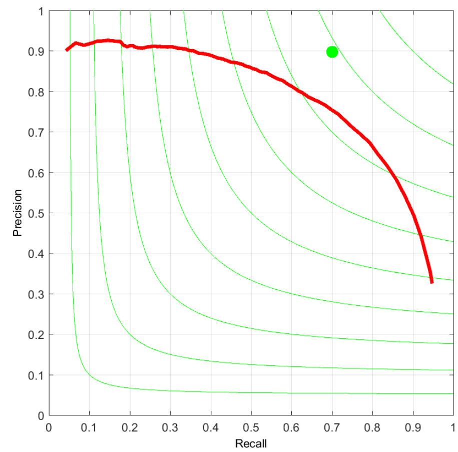

Figure 3: PR Curve for the UCM segmentation method

4. Discussion

For the methods implemented in this practice, the best outcome was that of GMM, its recall decays in an almost linear fashion with K, after an optimum (possibly the true number of partitions in the image), theres almost no increase in precision (Fig. 1). Furthermore, the Hierarchical Clustering method has an estrange behavior, its Precision and Recall grow linearly, at least in the K evaluated, even so its overall performance is poor compared to GMM.

The best performing method is UCM, a possible reason for the enormous performance gap between UCM and our methods is the fact that UCM does not establish a fixed number of partitions (K) a priori, it also takes into account multiple feature spaces related to image segmentation as texture or intensity gradients, compared to the simple color representation used for GMM and Hierarchical Clustering.

The results from the BSDS benchmark and the evaluation metric proposed in Practice 6 differ a lot, the PR curves show that GMM is clearly a superior method to Hierarchical Clustering although the similarity calculated with our metric was higher for the test case with hierarchical. Although a direct comparison between the two metrics wouldn't be justified, as in this practice a numerical threshold (or parameter) was used to construct the PR curve, compared to a fixed outcome for which our proposed metric was calculated.

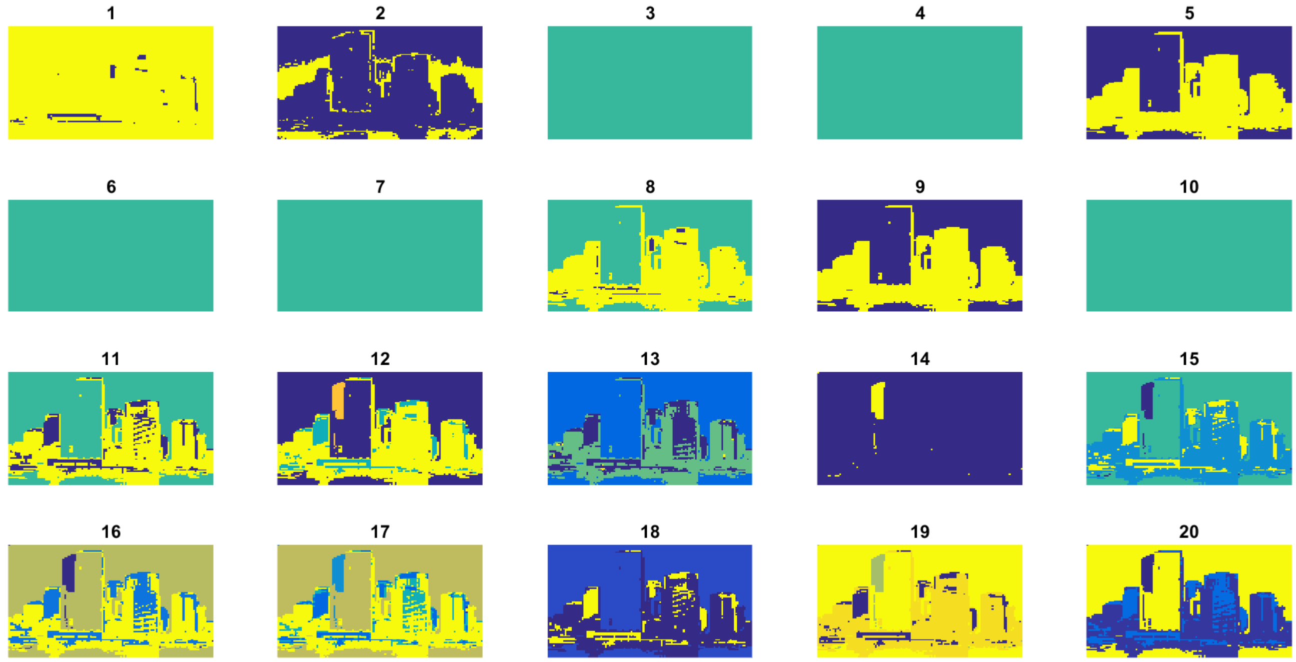

Figure 4: Multiple segmentation results for GMM

From the segmentation results in Fig. 4 its possible to observe that a clustering method like GMM does not recognize object boundaries (shadows possibly) as being part of the same object these delimit, effectively classifying borders as their own region. This error is persistent across all the number of partitions evaluated.

To improve our method is necessary to take into account other feature spaces, a problem with the traditional clustering methods is that they fit possible regions in a normal distribution, so its not possible to incorporate spatial information without fragmenting the image. Non-linear clustering methods like Mean-shift would be more appropriate for segmentation or, if the EM algorithm is used, to know the underlying distribution for each class of natural images (possibly not a Gaussian one).

References

[1] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. Contour detection and hierarchical image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 33(5):898–916, 2011.

[2] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proc. 8th Int'l Conf. Computer Vision*, volume 2, pages 416–423, July 2001.