Segmentation by clustering

Alejandro Trujillo
Universidad de los Andes
Bogota, Colombia
af.trujilloa@uniandes.edu.co

Santiago Martnez
Universidad de los Andes
Bogota, Colombia
s.martinez1@uniandes.edu.co

Abstract

In this report, we intent to use clustering methods in order to segment images of all kinds of classes. We tried 4 different clustering methods and 6 different feature spaces and tried to optimize our results. The best result we got by using our own evaluation method was a Gaussian Mixture model using a k of 7, However human segmentations resulted more complex than simple clustering algorithms, hich implies that we need to use a different aproach in order to improve our results.

In this laboratory we will use kmeans, herarchic gmm, watershed algorithms for the segmentation, each responds in a different way, what is intended is to observe the response with different levels of cluster that is givenfor us, the most complicated segmentation was that of watershed that when looking for a minimum with gradients, limiting it to the cluster number that was wanted was complicated, the database to use are 28 images that belonged to a larger file, that present a .mat file that will serve us to compare our results

1. Introduction

The problem of segmentation is a widely studied area when it comes to computer vision, and the formulation of the problem has varied over the years. During this laboratory, we will use clustering methods in order to divide the image into different regions that we hope represent semantic structures defined by a human observer. There are multiple algorithms to cluster data, the ones we'll use here are K-means, Gaussian Mixture Models (GMM), Hierarchical and watershed. Each one of these methods has it's advantages and limitations when it comes to image segmentation.

The K-means algorithm is a famous clustering method that, given a pre-defined number of clusters, divides the data comparing numerical information and minimizing the distance between each point of the cluster and its distance to the centroid of the cluster. It's an iterative algorithm that does not necessarily find the global minimum of the energy function that defines the problem, as it starts with randomized centroids and stops iterating when there's no a significant change in the centroid's position. Due to its iterative nature, it's an algorithm that requires a considerable amount of memory and processing power.

similar to the K-means algorithm, the GMM consists of adjusting an arbitrary number of gaussian distributions to the set of points in the selected feature space. The main advantage of GMM in comparison to the K-means algorithm is the possibility of clustering in different shapes, as K-means uses the euclidean distance which only permits circular clusters, while the Gaussian model is more flexible as it can take different shapes. Moreover, the GMM gives a "smooth" answer, as it's not a deterministic clustering, but gives a probability map of belonging to a given cluster, which gives more flexibility to the algorithm.

The Hierarchical clustering is an intuitive method for clustering. It consists of simply separating data by similarity on levels. Depending on how the algorithm is designed, the first level is either all data in one big cluster or every piece of data being its own cluster, while all the levels in between represent similar data clustered. This can be easily visualized in a dendogram. The parameter here is the amount of clusters we want, which translates into a level in the pyramid when it comes to the model.

The watershed algorithm is one that uses morphological mathematics to separate the image in regions. In this images it's a way to separate texture images in regions of interest, is a method with base in regions where we compared their gradient, textures, and spatial proximity,

2. Methodology

The Database used consists of a smaller version of the BSDS segmentation database. It was divided in a train folder (60 images) and validation folder (28 images), both

annotated by up to 6 different human segmentations. All images in this database have either a landscape orientation (481x321) or Portrait orientation (321x421).
In this experiment, we used six different feature spaces in order to segment an image using clustering algorithms. This spaces consisted of three color spaces (rgb, hsv and lab) and these same spaces but also including spatial information. The designed function takes in the ”featureSpace” parameter depending on what the user wants to use. The clustering algorithms used were k-means, Gaussian Mixture Model(GMM), a hierarchical model and watershed. For the first three methods, the descriptor used was a vector that included each component of the chromatic information (rgb, hsv or lab channels) and in case of specifying the use of spatial information, include the position of the pixel in the vector. As the dynamic range of each channel varies depending on the color space and, in case of the spatial information, the size of the image, all the information was normalized and forced into a dynamic range of 0-255, so each component carries the same importance for the algorithm. In the case of watershed segmentation, the magnitude of the gradient in the image was computed and the watershed algorithm was applied on it.

For the K-means and GMM methods,the image is rescaled to half of its size, in the case of the hierarchical, its scaled to a quarter of its size and for watershed, it wasn't rescaled. This resizing is done in order to reduce the time the algorithm takes to cluster the data and save memory. Although this process make the edges less smooth, the error isn't significant compared to the amount of time saved.

The method of evaluation used during this experiments was one made by ourselves. It consists of, separating the segmentation into binary masks and doing the same for the ground truth. After that, every layer of the new volume for the segmentations was compared to every layer of the ground truth. This was done by multiplying both layers and summing all the ones in the result, which would be equivalent to the intersection of the matrices and the dividing it by the sum of the matrix resulting from summing both matrices, which would be equivalent to the union. After comparing it to all other images, we take the one with the highest index value and assume it as the as the semantic structure in the human annotation. this was done for all structures in the computed segmentation and the mean of these numbers is the performance index we used. Applying this to the train image 24063.jpg, using K-means and 3 clusters we observed the following behaviour:

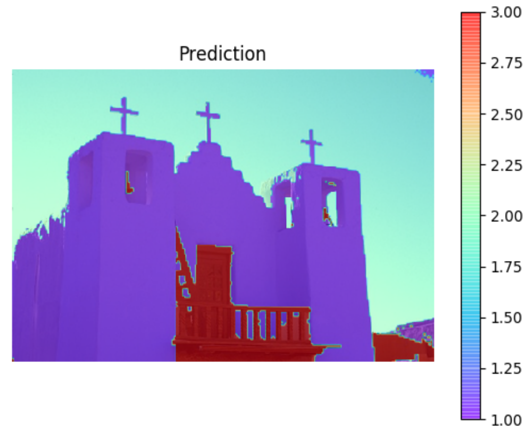

Figure 1. Computed segmentation overlapped with the original image

Figure 2. Ground Truth overlapped with the original image

And aplying the described evaluation method we obtained the following vector which represents how similar is the structure to its most similar counterpart in the ground truth: [0.0025777941447247283, 0.9609430195054524, 0.0], so our index of similarity is 0.321173604550059 for this k.

3. Results

This process was done for every image in the validation folder and the index was computed, all the indexes were summed and then divided by the amount of images in validation. The results were the following:

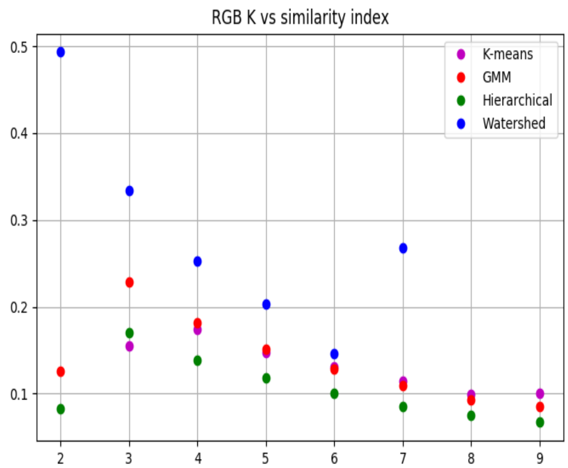

Figure 3. Performance for clustering using RGB

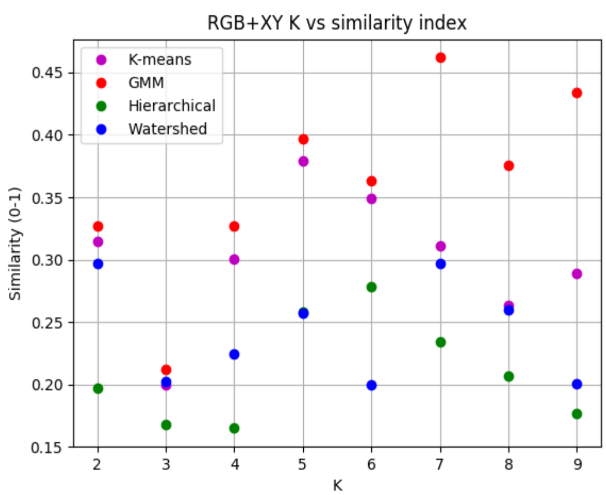

Figure 4. Performance for clustering using RGB and spatial information

We realized that images that are very homogeneous color-wise were the ones with the worst index. This is due to the limitations of the methods we are using, as all of them are based on the difference of pixels based on color or position, and having an homogeneous image the algorithm couldn't divide the structures as humans did. The evaluation strategy used here is biased towards lower numbers of clusters, as the computation of the mean ignores the fact that the algorithm may be ignoring or adding clusters, so a way to improve the evaluation would be to take into account the amount of clusters made by the human.

We know our evaluation method is biased towards small ks, so, although watersheds have a really high index in the RGB only segmentation, we know this result isn't actually the best. On the other hand, we know GMM had a relatively high similarity index (around 0.45) on the validation data with a k of 7, so we expect this to be our best parameter due to the optimization.

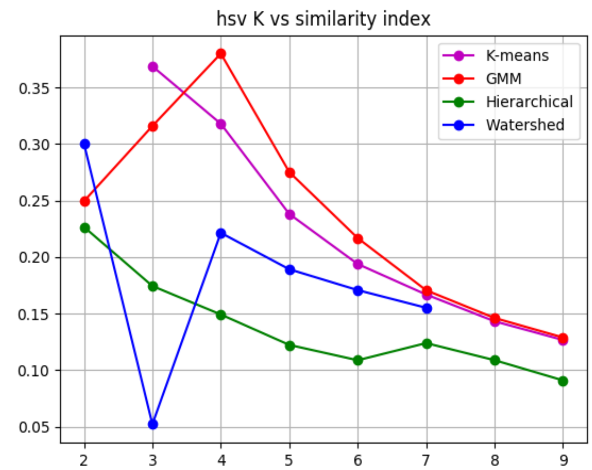

Figure 5. Computed segmentation overlapped with the original image

4. Conclusion

Even after trying multiple methods, the segmentation results weren't good enough compared to the human segmentation. This might imply that humans have top-down kind of processing which cannot be modeled under an unsupervised model. This means we need to use a different aproach for this problem if we want to improve our results. However for non-homogenous images, the GMM model worked decently for very simple applications.

References