Comparative Study between three segmentation algorithms with the Berkeley Segmentation Benchmark

Laura Munar Acosta
Universidad de los Andes
l.munar10@uniandes.edu.co

Maria Ana Ortiz
Universidad de los Andes
ma.ortiz1@uniandes.edu.co

1. Introduction

The image segmentation is referred to as one of the most important processes of image processing. Image segmentation technique is used to partition an image into meaningful parts having similar features and properties. The goal of image segmentation is to divide an image into several parts/segments having similar features or attributes. The basic applications of this problem are: content based image retrieval, medical imaging, object detection and recognition tasks [6]. The most popular techniques used for image segmentation are: thresholding method, edge detection based techniques, region based techniques, clustering based techniques, watershed based techniques and artificial neural network techniques.

The clustering based techniques are the techniques, which segment the image into clusters having pixels with similar characteristics. Data clustering is the method that divides the data elements into clusters such that elements in same cluster are more similar to each other than others [6]. One of the most used clustering algorithm is k-means clustering. It is simple and computationally faster than the hierarchical clustering. This algorithm aims to minimize the sum of squared distances between all points and the cluster centre. The main disadvantage of k-means is that produces different cluster results for each number of clusters; also, different value of initial centroid would result in a different cluster. So selection of proper initial centroid is an important task [5]. On the other hand, a Gaussian Mixture Model (GMMs) is a parametric probability density function represented as a weighted sum of Gaussian component densities. GMMs are commonly used as a parametric model of the probability distribution of continuous measurements or features in a biometric system. Its parameters are estimated from training data using the iterative Expectation-Maximization (EM) algorithm[9].

In the present work, the previous two segmentation algorithms are going to be compared versus the Ultrametic Contour Map (UCM). The later is an algorithm that produces a hierarchical segmentation from the output of any contour detector. This algorithm uses Oriented Watershed Transform (OWT) for producing a set of initial regions from contour detector output. Then, the UCM is constructed from the boundaries of these initial regions [2]. It is important to notice that Kmeans and GMM methods were selected because in a previous work they had the best performance between other segmentation techniques. Also, each algorithm is performed with a selected color space.

In order to understand the metrics for the comparison between algorithms, it is important to know the following. Precision is a ratio of the number of true positives divided by the sum of the true positives and false positives. It describes how good a model is at predicting the positive class. Precision is referred to as the positive predictive value. Recall is calculated as the ratio of the number of true positives divided by the sum of the true positives and the false negatives. Recall is the same as sensitivity. The precision-recall curve shows the tradeoff between precision and recall for different threshold. The Fmax calculates the harmonic mean of the precision and recall. According to that metric, the ideal result is to have 100% of precision and recall. [4]. Also, the Optimal Dataset Scale (ODS) is the best F-measure on the dataset for a fixed scale and the Optimal Image Scale (OIS) is the aggregate F-measure on the dataset fot the best scale in each image [2]. Finally, all these metrics can provide a point for comparison between the three methods in this study.

2. Materials and Methods

2.1. Methods

Methods used for segmentation were kmeans and mixture of gaussians, in HSV color space and LAB color space for a range of 0 to 20 clusters per image or 0 to 38 clusters per image. More specific kmeans was evaluated in HSV color space for 38 clusters and mixture of gaussians was evaluated in LAB color space for a range of 0 to 20 clusters.

2.2. Database

The ***BSDS500*** dataset contains 500 images from nature, containing groundtruth annotations made by an average of 5 humans. Dataset is split into train, validation and test subsets, train and test have 200 images and validation has 100 images[1]. Each image on the database can gave two size either $321x481$ (Portrait) or $481x321$ (Landscape),
BSDS500 evaluation method consist on evaluating problem as contour and region segmentation, meaning not only region intersection will be taken into account but that the contours of the objects matched. More specifically precision and recall curves were calculated by considering unmatched pixels as false positives and matched as positive, and allowing small localization errors by correspondence. Additionally as the dataset contains from 5-10 groudtruth annotations all where taken into account as adding them all together. [7], [3].
Sample Images From database A sample of the images and the annotations are shown in the figure 2 where each object has a colored label on top of the original image. Additionally on figure 1 an example of original images can be observed.

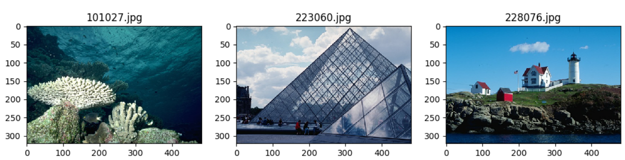

Figure 1. Original images for segmentation (3 from 200 available, title corresponds to file name)

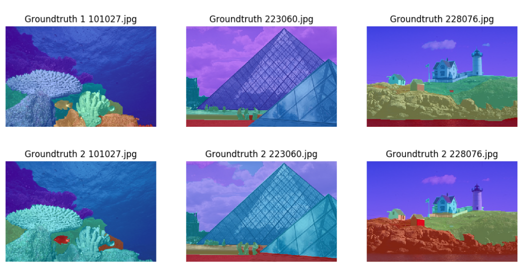

Figure 2. Groundtruth from 3 images and two different human segmentations

3. Results

The results obtained for the segmentation methods for an specific number of clusters are presented below (Figure 3): The precision-recall curve for the three segmentation methods is presented. For the curve presented of figure 4 it can be observed how the blue curves have a less parabolic

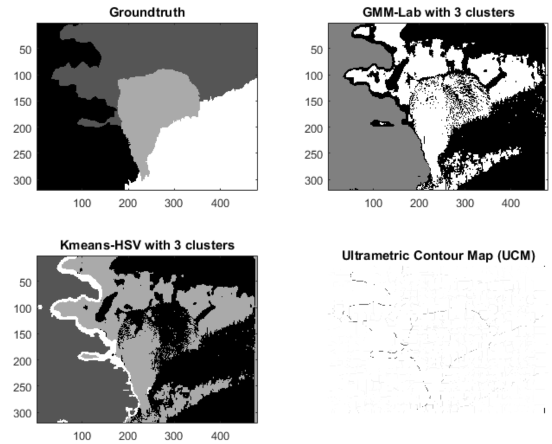

Figure 3. Segmentation results for the three methods

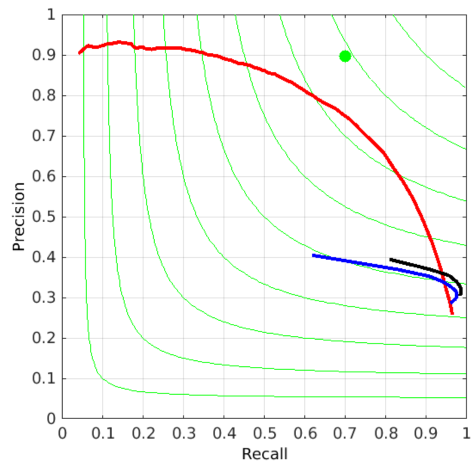

Figure 4. Precision-Recall curve for the methods: UCM2 (in red), GMM-Lab (in blue) and Kmeans-HSV (in black)

shape. Additionally blue curves does not covers the whole graph.

Now, the metrics obtained for the methods are shown in Table 1.

On figure 5 it can be observed the segmentations obtained for mixture of gaussians method on different amount of clusters, on this result it can be observed a great variety per each number of clusters segmentated.

Finally for figure 6 it can be observed KMeans on different number of clusters, on 3 clusters it can be observed a good result, however as clusters grow less objects can be observed.

4. Discussion

As seen in Figure 3, the GMM-Lab and Kmeans-HSV methods do not segmentate the picture in the regions

Table 1. Segmentation metrics for test data

	Covering			
	ODS	**th/K cluster**	**OIS**	**Best**
UCM	0.59	0.2	0.65	0.74
GMM-Lab	0.28	4	0.36	-
Kmeans-HSV	0.4	2	0.43	-

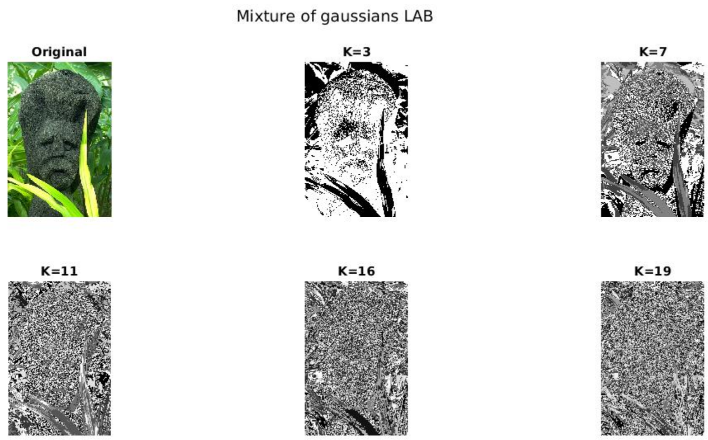

Figure 5. Method: Mixture of Gaussians for LAB color space for different amount of clusters

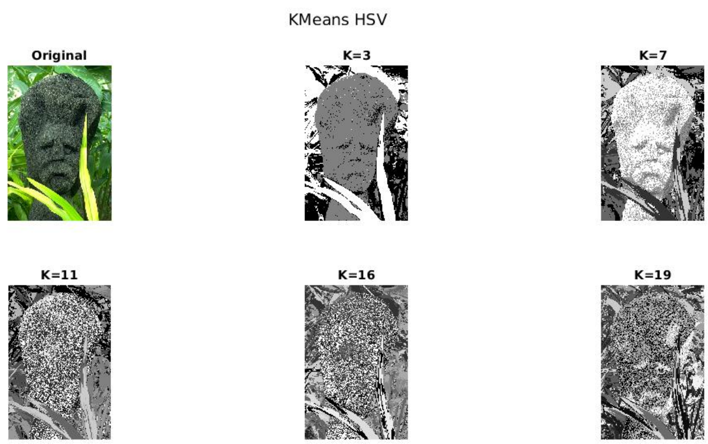

Figure 6. Method: KMeans for HSV color space for different amount of clusters

wanted. These methods had as main feature the color space and it seems that the picture has predominant textures so the algorithms can not recognize well the contours. On the other hand, the UCM presents a hierarchical information of the segmentation but in general it can be seen that the method can separate the parts as in the groundtruth. But, there is a need to obtain segmentations on this method by thresholding at the optimal dataset scale (ODS) the UCM. The main difference on the performance of these method, as mentioned before, is that Ultrametric Contour Map (UCM) defines a duality between closed, nonself-intersecting weighted contours and a hierarchy of regions. Making this shift in representation from a single segmentation to a nested collection of segmentations turns out to be very powerful. So, UCMs combines several types of low level image information into a generic notion of segmentation scale, constructing a hierarchical representation [2]. Conversely, Kmeans and GMM does not have the hierarchical information so they can not discriminate if it is an important information within all levels or not.

In the table 1 is shown the metrics results for the algorithm. Where it can be seen that the UCM method is better in the metric presented (Covering). In the Optimal Dataset Scale (ODS) UCM presents values in 0.59 in covering, which is the overall Jaccard Index for the entire dataset. In contrast, GMM-Lab and Kmeans-HSV obtained 0.28 and 0.4 respectively. In this table, we can also see that the K number of clusters for these last methods are 4 and 2 in their best performance. Finally, it is seen that the OIS is slightly greater than ODS and this is expected because this metric represents the performance per image and not between the entire dataset.

Furthermore, the Figure 4 shows that the UCM method presents a greater area under the curve than Kmeans and GMM. There is a big difference between the precision of the three methods. Here we can see the importance of considering hierarchical information, because this gives information between levels and preserves the contours that predominated in them. This figure also shows that the recall for Kmeans and GMM is more dominant than the precision. That means that for a high recall the precision of these methods is low. So, these methods are not good distinguising between classes and they are not precise in the segmentation.

Errors observed on segmentation for the two methods follow a pattern as observed of figures 5 and 6 in which when K gets bigger less objects can be observed due to over-segmentation, this can be observed mainly because it uses only color information and as textures varied on color, it make it hard to segmentate objects with similar textures.

4.1. Best Method

From previous report the best method was watersheds and KMeans, however for this report best methods evaluated where KMeans and Mixture of Gaussians. So by comparing performance KMeans on HSV space was still the best one, nonetheless performance was significantly apart from UCM method [8]. UCM method was the best because it takes into consideration more representations of the images, and uses hierarchical contours, in parellel KMeans and Mixture of gaussians ignores important information that may came from different characteristics as textures.

4.2. Methods Limitations

Limitations to the methods used are mostly related to the input information, mainly because local information is not characterize enough allowing to miss objects that may varied on color but are similar due to its local information or to join objects far away but that has the same color. Regarding what was previously mentioned it was not enough for the methods to add additional dimensions regarding position. Another limitations identify was that methods vary on founding boundary objects, mainly because it takes similarity on colors because that was the characteristic used to group. Finally as briefly mentioned before to cluster the image based on its color similarities was the main limitation, because color misses position information and object instance information.

5. Conclusions

After evaluating results and discussion it can be concluded that color information for clustering methods is not enough because as mention previously instance of objects share more than same color characteristic. Additionally methods proposed can be improved by adding more characteristics to pixels, such as neighbors information, and more specific position information, and its relations to their neighbors. Finally, it can be concluded that characterization of the image is very important, as observed for UCM method that uses several characteristics to get to the probability of boundary.

References

[1] UC Berkeley Computer Vision Group - Contour Detection and Image Segmentation - Resources.

[2] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. From contours to regions: An empirical evaluation. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2294–2301. IEEE, 2009.

[3] 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, May 2011.

[4] J. Brownlee. How and when to use roc curves and precision-recall curves for classification in python. Available: https://machinelearningmastery.com/roc-curves-and-precision-recall-curves-for-classification-in-python/.

[5] N. Dhanachandra, K. Manglem, and Y. J. Chanu. Image segmentation using k-means clustering algorithm and subtractive clustering algorithm. *Procedia Computer Science*, 54:764–771, 2015.

[6] D. Kaur and Y. Kaur. Various image segmentation techniques: a review. *International Journal of Computer Science and Mobile Computing*, 3(5):809–814, 2014.

[7] D. R. Martin, C. C. Fowlkes, and J. Malik. Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(5):530–549, May 2004.

[8] L. Munar and M. Ortiz. Segmentation using clustering methods. *Report Laboratory for Artificial Vision*, 2019.

[9] H. Permuter, J. Francos, and I. Jermyn. A study of gaussian mixture models of color and texture features for image classification and segmentation. *Pattern Recognition*, 39(4):695–706, 2006.