Evaluation of Unsupervised Clustering Segmentation Methods

Luisa Fernanda Borbón R.
Universidad de los Andes
Departamento de Ingeniera Biomdica
lf.borbon@uniandes.edu.co

Abstract

Segmentation is an image processing technique that aims to divide an image into its relevant objects, according to similar features like color, texture and position. In this sense, throughout this work the problem of segmenting the images from BSDS500 dataset was approached by using K-means, GMM and UCM algorithm. The best algorithm obtained was not able to get results as well as the gPb-OWT-UCM algorithm. In order to improve the method presented, more local information as texture or borders should be use to segment the desired objects.

1. Introduction

Image Segmentation is an important and challenging technique of image processing, basis of recognition problems resolution and image semantic characterization. This technique is used to break down an image into its meaningful parts, by dividing it according to similar local features like color, texture, brightness or spatial position [4]. For this task, clustering algorithms such as K-means and Gaussian Mixture Models have been used, specially when there is no previous knowledge of the desired objects to segment [2].

Regarding to the segmentation problem, one of the main universities that worked in this topic was the Berkeley University of California, with its Computer Vision Group. In 2001, this group released the Berkeley Segmentation Dataset, with 1000 RGB images of resolution 481x321 taken from the Corel image database. Each image had at least one discernible object and counted with different human annotations. The availability of this dataset allowed researchers to join the segmentation problems area an contribute improving faster algorithms than the ones used at the time. In addition to this, as the BSDS database was set to public with annotations, new evaluation criteria to compare the results of segmentation algorithms like the precision and recall curve were created [1].

Due to the importance of segmentation problems in image computer processing, the present work intents to compare the segmentation results of three methods in the BSDS500 dataset, using the precision and recall curve, maximum F measure and area below the PR curve. The methods that are going to be compared are: K-Means, Gaussian Mixture Model (GMM) and UCM; and also the color space representation of the images is going to be changed between lab+xy and rgb+xy, in order to find the algorithm with the best performance.

2. Materials and Methods

To compare the different segmentation algorithms, the BSDS500 dataset was used. The kmeans and gmm methods were implemented by using the python segmentation function created in a previous work and the ucm function was taken from Berkeley Computer Vision Group's open source algorithm [6]. Below, a detailed description of the Dataset, segmentation methods and evaluation methodology is found.

2.1. Dataset

The BSDS500 image set is made up by 500 natural images with ground-truth human annotations and benchmarking code. The dataset is distributed in 200 train and test images and 100 validation instances, and all the images come in *.jpg* format, have at least 3 human made segmentations and a size of 481x321 pixels both vertically and horizontally oriented [3]. In addition, this collection presents a big variety of scenes and objects as shows the test image examples in figure 1. This variability increases the difficulty of segmentation, as it makes the instances of each image difficult to characterize with an steady parameter model.

2.2. Segmentation Methods

As mentioned before, in order to segment the images with the methods of k-means and gmm, the python function segmentByClustering was used. The K-Means algorithm uses the euclidean distance to group all the data near to previously defined randomly or centroids, and optimizes its

Figure 1. BSDS500 test image examples.

position until is stabilizes according to the other points. On the other hand, the Gaussian Mixture Model or GMM fits the image data as a mixture of Gaussians with k clusters, giving a soft assignment or probability of each point to fit inside each Gaussian distribution.

In the previous work, using this function and the proposed evaluation metric, the best methods were the Gaussian Mixture Model in the rgb+xy color space, followed by K-Means clustering using the lab+xy color space. As in both cases, the spatial position of the pixels seemed to give valuable information, for this test both color spaces (lab+xy and rgb+xy) and methods (Kmeans and Gmm) were used.

The only parameter mandatory to use the previously mentioned function is the number of clusters wanted (k). This parameter is chosen according to the number of objects to segment in the image, but it also depends on the computational resources available as using a larger k number requires more processing time and power. In this case, the k parameter was varied in order to obtain the precision and recall curve and be able to compare the different methods trough it and the other evaluation metrics. The number of clusters used on each train, validation and test image to evaluate the algorithm, varied between 5 and 80 with steps of 15, for a total of 5 segmentations per image. Despite not being a lot of segmentations, this wide range was selected to ensure a rich representation of and a faster computational process.

2.3. Evaluation

Lastly, the umc method and the evaluation metric comparison was done with the image segmentation resources from Berkeley's Computer Vision Group, specifically the allBench-Fast( ) matlab function. This resources help with the evaluation method by giving boundary and region metrics results for each segmentation method. The principal way to evaluate this dataset is the Precision and Recall curve, constructed as mentioned before by varying the number of clusters and calculating the accuracy in precision and recall of the method. Other important metrics used were the Area under the PR curve, the F-Measure and two metrics obtained from it: Optimal data Scale (ODS) and Optima image scale (OIS). Since this is a problem of segmentation, the Segmentation Covering(SC) degree of overlap between a segmentation an its corresponding groundtruth, the Probabilistic Rand Index (PRI) the fraction of pixels labelled correctly and variation of information (VI) or the distance between segmentations in terms of mutual information or entropy, were also taken into account in order to decide the best segmentation algorithm. [5]

3. Results

The BSDS500 images were segmented using the clustering methods K-Means and GMM, with both rgb+xy and lab+xy color spaces. The Precision and Recall curve associated to the algorithms was created using 5 segmentations per image and varying between 5 and 80 with steps of 15, as mentioned before. The algorithm was first runned with the Gaussian Mixture Model method in both color spaces, the results of the gPb-OWT-UCM algorithm and is shown in figure 2.

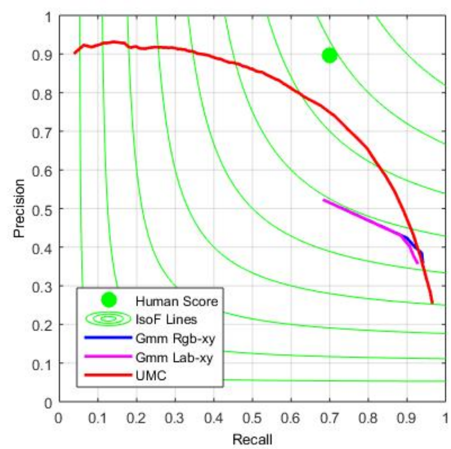

Figure 2. PR curve for the BSDS500 test using the ucm method and gmm with lab+xy and rgb+xy color spaces.

Figure 2 shows that even though varying the color space between lab+xy and rgb+xy when using k-means doesn't have a big impact on the PR curve, the method with lab color space presents a wider range of action, being able to segment with higher precision as the recall goes down. The same construction of the PR curve was followed with the k-means method, and its resulting graph can be seen in figure 3. This figure shows a consistent results when comparing to the previous method, in which using the method using the lab+xy space has a wider range of activity and may present a better segmentation algorithm when applied to specific cases. As the best methods were the

Table 1. Quantitative Metrics evaluated for the segmentation algorithms Gmm, K-means and ucm.

	Gmm		K-means		
	Lab+xy	Rgb-xy	Lab-xy	Rgb-xy	ucm
ODS	0.58	0.57	0.57	0.56	0.73
OIS	0.63	0.63	0.60	0.60	0.76
Area	0.12	0.1	0.27	0.13	0.73

ones that used the lab+xy color space, the figure 4 shows the PR curve for ucm, gmm and k-means, this last two using the lab+xy color representation.

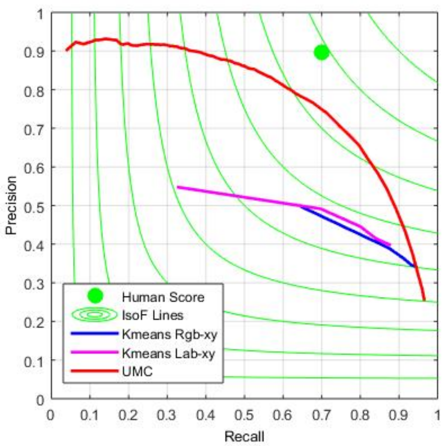

Figure 3. PR curve for the BSDS500 test using the ucm method and k-means with lab+xy and rgb+xy color spaces.

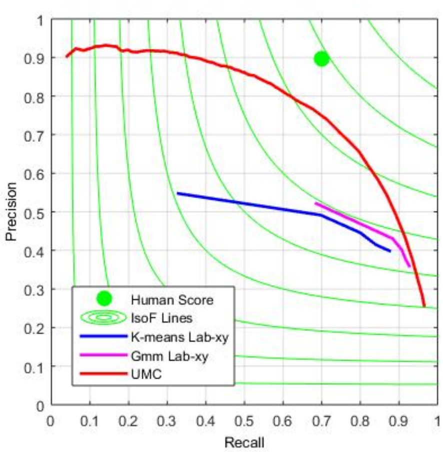

Figure 4. Watershed Clustering algorithm applied with k=5. The color space used and the metric number evaluation is shown on top of each result.

In order to define which method is better, the segmentations associated to these methods were visually inspected as shown in figure 5 and the other metrics were calculated and shown in table 1.

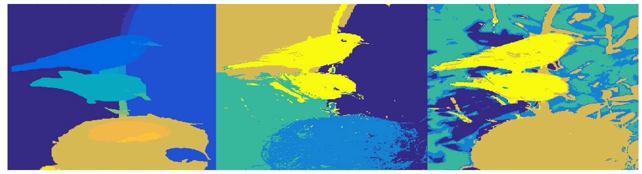

Figure 5. Examples of the best

4. Discussion

Comparing all the PR curves obtained and taking into account the results exposed, the algorithm with the best performance is Gmm using the color space lab+xy. This could be due to the soft assignment done by this method, that presents itself as a better option compared to the hard assign done by k-means.

The position representation shown to be valuable when segmenting an image and the fact that lab is an uniform space color may also be helping to create a good representation of the image; but even if this algorithm displays the best results, the gPb-OWT-UCM algorithm is way better. This is probably caused by the fact that the gPb-OWT-UCM uses a wider range of representation that involves more local infomation given from texture, brightness and color.

The gmm model assumes a gaussian distribution of the image information this is not usually the case. To improve the method created, the local characteristics should be studied in further detail, changing more parameters on the clustering methods in order to adjust better to the dataset proposed. Following this idea, the inclusion of a wider range of local descriptors should also be involved to improve the method. As the xy channels for position were added, the same could be done with texture of borders to obtain a better characterization.

Other important aspect to take into account are the metrics used for comparison. As well as the segmentation metric proposed in the last work, they are not all consistent as the measure and take into account different information. So its important to select a metric according to the applications or the expected results of the algorithms; either if a higher precision or recall are needed, or any other specifications regarding to the problem that is trying to be solved.

5. Conclusions

The BSDS500 dataset presented itself as an innovative dataset that lead into new segmentation algorithms and evaluation methods. For segmentation, clustering methods seemed as a good approach but in reality are not able to get to the level of actual famous and more complex state of the art methods as gPb-OWT-UCM. Again, the space of representation of the images and its evaluation metrics played a main role, as the have to be adjusted depending on the problem that is trying to be solved.

References

[1] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. Contour detection and hierarchical image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(5):898–916, May 2011.

[2] P. Arbelez. Lecture 5: Clustering, computer vision, 2019. Universidad de los Andes.

[3] Berkeley. Contour detection and image segmentation resources, 2011. [Online] https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/resources.htmlalgorithms.

[4] K. Dilpreet. Various image segmentation techniques: A review. International Journal of Computer Science and Mobile Computing, Vol. 3, Issue. 5, May 2014, pg.809 to 814.

[5] V. Ferrari. Computer vision- eccv 2018, 2018. 15th European Conference, Munich, Germany, Proceedings, Part III.

[6] 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. 2:416–423, July 2001.