Unsupervised Segmentation Methods

Alejandro Pardo
Universidad de los Andes
la.pardo2014@uniandes.edu.co

Lina Barbosa
Universidad de los Andes
lm.barbosa1099@uniandes.edu.co

Abstract

We propose four unsupervised methods for image segmentation: Kmeans, Mixure of Gaussians, Watershed and Hierchical Methods based on pixels color and position. All of them were evaluated using the Berkeley Segmentation Database (BSDS500)[2], in which the highest performance was obtained by Kmeans in lab+xy space and Watershed methods. The selected methods were compared with the Ultrametric Contour Map (UCM) algorithm.

1. Introduction

Image segementation is one of the most studied problems in Computer Vision. The goal of this kind of algorithms is to extract objects from images, but it is not as easy at it sounds. In last years, there have been developed lots of algorithms in order to improve the performance of the segmentations and in that way getting closer to the one of humans.Ultrametric Contour Map (UCM) is one of the algorithms with best score of performance and is part of the benchmark of Berkeley Segmentation Dataset (BSDS500) to eavluate segmentation algorithms. In this work we propose four methods for color image segmentation: Kmeans, Mixure of Gaussians, Watersheds and Hierachical clustering based on color space and position. All of our methods were evaluated using BSDS500 and finally compared with the segmentation results for UCM algorithm in order to obtain an score for its performance and relate it with the state of art.

2. Methods

2.1. General Description

We proposed four unsupervised segmentation methods(without previous learning of the dataset). The image segmentation was based on image color space and position for each of the points. The color spaces employed were RGB, LAB and HSV. In this way, each pixel in the image will have an own descriptor including color space intensities and position; for example a descriptor D could be defined in this way: D=(R,G,B,X,Y). Methodsddddd will be described below:

2.1.1 K-Means

This method need a specific space of representation with n dimentions that are given by the experimenter and depends on the problem to approach. After setting the space of representation, every candidate (in this case, every pixel) is represented by a descriptor, which is a vector with dimension n which contains a value for every dimension of the space. Then, every candidate becomes in a point on the space. [1] After arrange all the points in the space, K-Means try to separate the space in k groups, based on the distance between points. K is previously known. Something important about this method is the final distribution of the groups, kmeans doesn't requires any probabilistic modeling, and implicitly models the probability density as spherical distributions.[1]

2.1.2 Mixture of Gaussians Method (GMM)

The global idea of this method is to represent each cluster with a Gaussian distribution, so the goal is to optimally fit the model to the data, this implies to estimate the optimal parameters if a mixture of Gasussians in order to explain the observed data. To associate input samples with cluster center the *Mahalanobis distance* is used instead of using nearest neighbors as Kmeans method: this allows distance to adapt to the data distribution. The approximate solution to this method is basen on Expectation Maximizatiuon algorithm (EM), in which the expectation (E) step estimate responsabilities given the model parameters and de maximization (M) step estimate paraeters, given these responsabilities. This gives the method a soft-assigmnment behaviour. It means that ir produces for an input responsabilities for each cluster. It results important to mention that the method requires prior knowledge of number of models, K, and that the approximate solution depends on the initializations and can converge to a local minimum. [1]

2.1.3 Watersheds

Watersheds technique find different regions of an image by findind the catching basis, which are segments that divide the local minima of the image. This segments, can be considered the limits of the regions, therefore could be a potential edge. The main problem of this method, is that it associates unique region with each local minimun, and this lead to over- segmentation problems. One alternative to solve this issue, is the h-minima, which force to the method to find only the minima that pass the threshold given by h. This reduce the amount of minimums finded and can give to the image a better aproximation of its regions. [1]

2.1.4 Hierarchical Clustering

The global idea o0f this type of methods is to define a distance D between groups based on a similarity meassure between elements. It creates a hierarchy based on distance D, which can be represented by a hierarchical tree or dendogram. The construction algorithm can be agglomerative or divisive. In this case we perform and agglomerative algorithm which is based on the idea of assuming each point as a cluster to then merge two nearest clusters iteratively until obtaining a single cluster.[1] This was made using Euclidean distance to join near clusters.

3. Database

To evaluate the proposed methods we used the Berkeley Segmentation Dataset and Benchamrk (BSDS500) which includes 500 natural color images with size 321x481 pixels and jpg file format. This images were divided into a training set of 200 images, a validation test of 100 images and a test set of 200 images.

All of the dataset was manually segmented by 5 subjects and this human annotations served as ground truth as well as a benckmark for comparing different segmentation algorithms as all color segmentations generated by different algorithms, in this case the Ultrametric Contour Map (UCM) algorithm. The goal of benchmarksis to produce a score for the algorithm's segmentation, so that the proposed method can be compared to each other. This score was obtained using the F measure, which is the harmonic mean of precision and recall.

An example of available groundtruth on database is shown in Fig. 1, in which it can observed the human annotations of two disfferent subjects.

4. Methodology

To achieve the comparisson between methods we followed the steps described below:

1. Training: in this stage we adjust our methods using the train set of database to obtain the best performance for each one. We made visual experiments, and a qualitative evaluation to improve the performance of each method. We adjusted parameters like the weight of every dimension in our descriptor. In the case of the descriptors that involves color spaces and pixels posicions we give to position a much lower weight, by dividing its values by a factor of eight, giving to the color space the main role in the descriptor. We also visualize which channels were more important, and we gave them more importance in the descriptor. We made this process for every space, and every method.

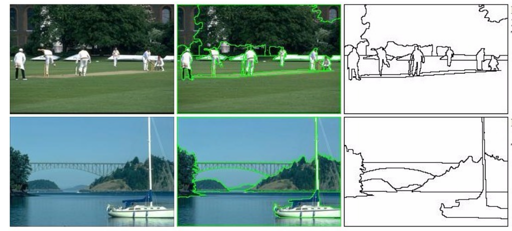

Figure 1. Example of Groundtruth based on annotations made by two different subjects

2. Validation: in this stage we probe our methods with the validation set of database. We generate a PR curve for each method and selected the two ones with higher F-measure. To ilustrate the results, we will show you one example of segmentation for a couple of images with the best methods:

Figure 2. Original input number 1, taken from the validation dataset of BSD500.

This is a really simple image, which we can see two regions, the plane and the sky, so a segmentation should be very simple.

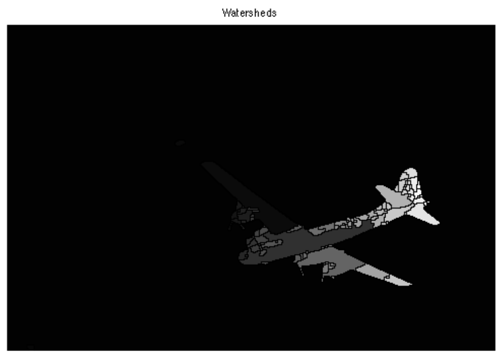

Figure 3. Watersheds segmentation of first image with h=1.

In this image we can see a segmentation with h=0.1, we can see that the plane is recovered by the method, however it is divided in several regions due to the local minima.

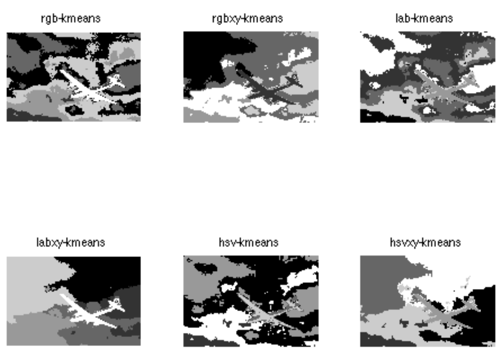

Figure 4. K-Means segmentation with k=3 and all the combinations of representation space.

In this case we can see the response for K-Means segmentation with k=3, we cans see in the image corresponding to Lab+xy a good segmentation of the plane. However, we oberserve a segmented background, this is the consecuence of the number of clusters defined (3 clusters), however we can appreciate a good segmentation in this particular case.

Aditionally, we'd like to show another example of the validation dataset, with a more complex image:

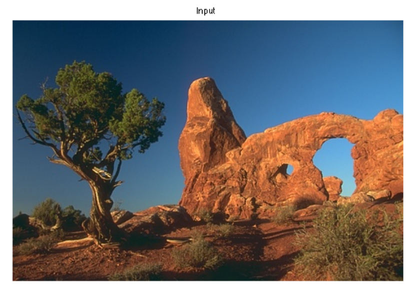

Figure 5. Original input number 1, taken from the validation dataset of BSD500.

As we can see, this is a more complex image with more objects and a big contrast of the backgorund that make it harder to segment.

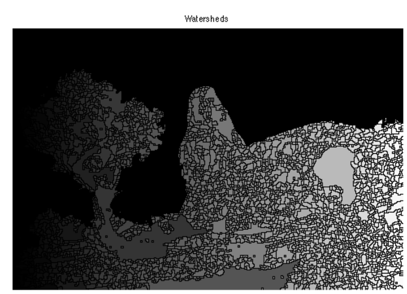

Figure 6. Watersheds segmentation of first image with h=1.

As expected, the result of Watersheds extracts the information of the objects and the background, but again we can see an over segmentation problem, as well as the first example.

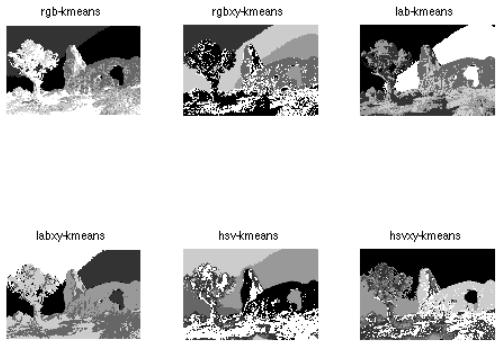

Figure 7. K-Means segmentation with k=3 and all the combinations of representation space.

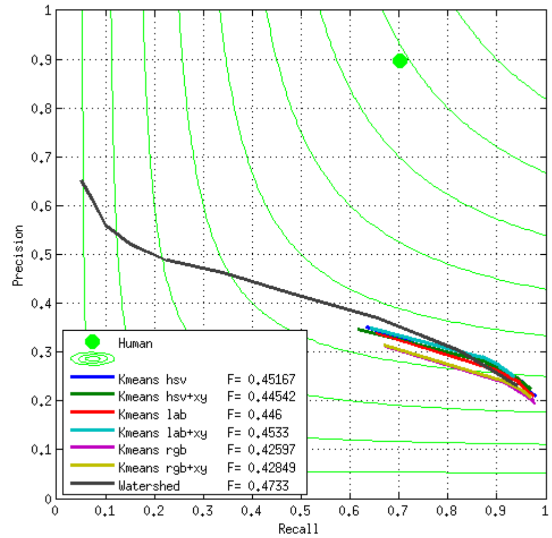

Figure 8. Methods Comparisson using validation set

K-Means shows again that the best representation space is Lab+xy, because with this space the segmentation is accomplished, we can see the three separated of the mountain and the sky. We also can observe that the responses of the other descriptor, are not as good as Lab+xy descriptor.

The GMM and hierarchical clustering methods were not evaluated because they were computationally expensive and presented difficulties to achieve convergency, during the training stage making ir harder to perform a posterior evaluation. In the case of hierarchical clustering, it was necessary to resize the input image to obtain appropiate results, but in this cases, there was lost of information on segmented image after returning to the original shape. On th other hand, GMM had lots of convergency problems in some of the images as the number of clusters where higher. Based on this reasons, we opted for excluding this methods from the validation and test sets in database.

In order to confirm the information obtained by these examples, we create a PR curve with the objective of selecting those two methods with the higher value of F- measure. The PR curves and the F- measure for each of the proposed methods is presented in Fig.8. We can see that the method with higher performance are watershed (F=0.47) and Kmeans using lab+xy space (F=0.45). This effectively confirm the behaviour described previously.

3. Test: Finally, using the test set of databse we evaluate the selected methods and compared them with the UCM algorithm. This was made through OR cruve and F-measure. The results of this item, can be seen at Results section.

5. Results

Based on the results obtained in the validation stage, we proceded to evaluated the selected methods in the test set of database in order to get an score of their performance and to make a comparisson with the UCM algorithm. Based on this, we create a PR curve which will let us to compare the three methods and calculate F-measure as a performance score. This curve is shown in Fig.9, we can see that the performance of the UCM algorithm is so much higher than the ones for our methods. On the second hand, it can be observed that the recall range for both methods y not as large as UCM's.

Method	F-Measure
UCM	0.72
Watershed	0.47
Kmeans LAB+XY	0.45

Table 1. F-measure obtained for evaluated methods

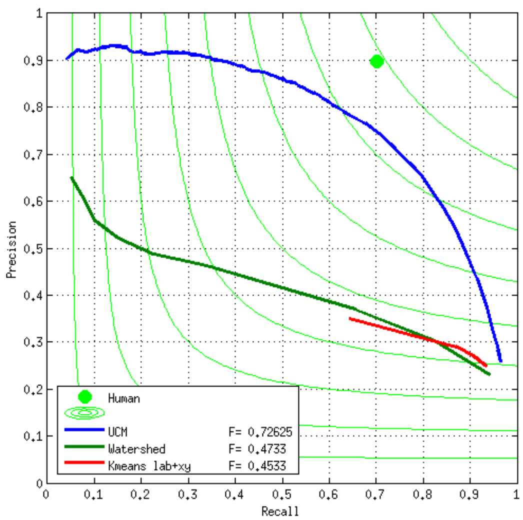

Figure 9. PR Curve comparing selected methods and UCM algorithm in test set

The F-measure obtained for the evaluated methods are summarized in Table 1, we can see that for both methods, the one for UCM algorithm is so much higher.

To give a couple of examples of the Resulst obtained, we will show you the segmentation for two images from Test dataset:

Figure 10. Original input number 1 for test, taken from the test dataset of BSD500.

This image is one of the hardest of the dataset, due to the textures of the objects and the holes that allow to see the backgorund trough the main object of the image.

Figure 11. Watersheds segmentation of first test image with h=1.

As expected, the result of Watersheds is an over-segmented result, that show us little information about the objects in the image.

Figure 12. K-Means with Lab+XY descriptor

We can see in this image a good result of the segmentation, because the objects are stracted from the background, however we can note that in the background, the method is over-segmentating, due to the distances, in the representation space, betwen pixels of the same object of the background.

The GroundTruth for this image is the following one:

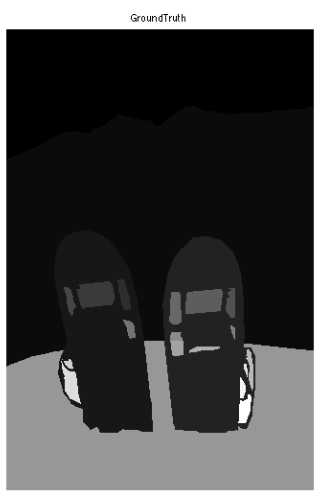

Figure 13. K-Means with Lab+XY descriptor

Here we can see the complexity of the image, with regions not well differentiated, and with a confusing GroundTruth, showing that even for a human it is a hard problem. This could be a reason of the problem of our methods to segment this particular image.

Another example is shown below:

Figure 14. Original input number 1 for test, taken from the test dataset of BSD500.

On the other hand, we can see this as a simple image whit only one object that would be the Polar Bear and a couple of elements in the background.

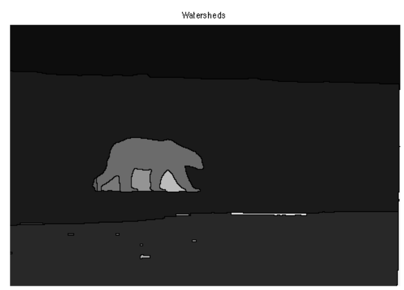

Figure 15. Watersheds segmentation of first test image with h=1.

Here watersheds haven't has a critical problem of over-segmentation and we can see clearly the components of the image: The Polar Bear, two shades of the snow, and the see.So in contrast with the groundTruth this could be a good segmentation.

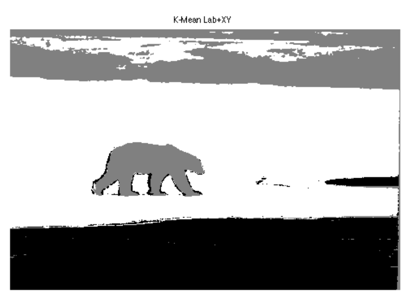

Figure 16. K-Means with Lab+XY descriptor

Finally, We see the result of kmean with Lab and XY posicions. It is a good segmentation based on the

GroundTruth images. We easily can see the different components of the image, without the presence of hard over-segmentation problems.

Now, see the groundTruth:

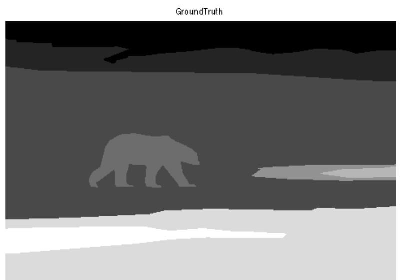

Figure 17. K-Means with Lab+XY descriptor

Here we can see that the image is easily segmented by the humans. Clearly differentiating the bear from the other objetcs of the background. We also can see the our methods are getting closer to this kind of result, with simple images like this.

6. Discussion

As we could see in the Results section, the algorithms implemented weren't even close to the behavior showed by the UCM algorithm. PR curve shows that UCM method is the closest to the human behavior, and our methods (Watersheds and K-Means with Lab space) were really far to reach this target point (0.8). We also can compare this results with the maximal F-measure, which shows us again that UCM is the best method, however between the couple of methods evaluated across this report, we can conclude that Watersheds could be a better segmentation method than K-Means. This can be seen by the range of the PR curve for watersehds, which shows us that maximum precision and maximum recall can be obtained through this method. The reason of this effect, is the number of clusters that we can obtain using watersheds, we can obtain a very large range of clusters with allows us to move through all the PR curve. On the other hand, K-Means allows a small number of clusters, due to te convergence of the method. Despite what we mentioned before, the maximal F-measure of both methods is really similar. This can be explained through the analysis of the chosen descriptor. We chosed a descriptor that only take care of the color space representation and positions of pixels on the image. However, it seems to not be enough, images have any others features that can be usen, like texture, relations between pixels and its neighbors, similarity between pixels of the image, etc. that could be really helpful to build a complete descriptor of the image. In the case of UCM, it uses much more complex descriptors that give to this method the great behavior showed in the BSD500. So, if we have to chose one of the methods, surely we would choose UCM mehthod, becuase of this behavior showed, and its constant behavior trough all the dataset. We already mentioned that the principal limitations of our methods were the descriptors. However, we haven't mentioned the limitation for UCM. Despite it is hard to think, becaause it presents a really similar behavior with humans, it is not exactly the same. Humans use other kind of information to segment an image, like functionality of the objects present in the images, and off course previous knowledge. On the other hand, UCM only use the information presented by the images, and it doesn't take care about previous experiences, so it have to be the greater limitation to this great method.

7. Improvements

We propose some imporvements to the algorithms that can lead to achieve a higher performance in image segmentations:

- The first one consist of extending the dimension of the descriptor for each pixel in the input image. It could be useful to add other properties such a texture responde over a neighborhood, afinity measurement with closest pixels or oriented gradients. This could improve the results because in this way, it will result easier to group pixels that belong to the same cluster.

- All of the methods evaluated used local characteristics to estimate affinity between pixel and in that way generate the clusters. Even though the local analysis could not be enough to extract objects from tan image. In this order of ideas, it will result usefull to add global descritors with the objective of analyzing the relation between each pixel and its environment inside the image. This coul be performed employing higher windows for the analysis.

- Supervised clasification might help a lot with boundary detection. Binary clasificators could be used to clasify every pixel of the image as a border/no border pixel. However, to approach to the segmentation problem, this kind of clasifcators, that use binary clasification might be not helpul, because categories in this case are larger and more diverse.

  incluir otras cosas, como textura, afinidad con otros pixeles, caracteristicas locales

References

[1] Szeliski, R. (2010). Computer vision: algorithms and applications. Springer Science & Business Media.

[2] Contour Detection and Hierarchical Image Segmentation P. Arbelaez, M. Maire, C. Fowlkes and J. Malik. IEEE TPAMI, Vol. 33, No. 5, pp. 898-916, May 2011.