Segmentation

Maria Camila Escobar
Universidad de los Andes
Bogotá D.C, Colombia
mc.escobar11@uniandes.edu.co

Laura Gongas
Universidad de los Andes
Bogotá D.C, Colombia
l.gongas10@uniandes.edu.co

***Abstract*- Image segmentation is a growing field in the computer vision market. Nowadays, with the design of self driving cars robots that can move through spaces we are in need of efficient pattern recognition methods and scene analysis. Both of this applications are based on a correct image segmentation. Here we develop a method for segmenting images with three different color spaces and four clustering methods. We create our own evaluation metric and evaluate the performance of our method in a fragment of the BSDS database. Finally, for each clustering method we analyze the outcomes and conclude about the overall performance of our method.**

1. Introduction

Image segmentation is generally essential as a preliminary step in scene analysis and automatic pictorial pattern recognition problems [1]. Segmentation algorithms aim to group pixels in homogeneous regions generally by a common feature approach between pixels. Features might be represented using space of color, texture and gray levels [4]. Also, the position of the pixel can be taken into account as a way to take into account spatial information.

Clustering is a classification method that works by representing each pixel with a vector of measurements, then it's compared with a similarity measurement to other pixel vectors and finally each pixel is labeled at a corresponding cluster. In other words, clustering is an indicator of similarity of image regions and therefore it's useful for segmentation. Also, it's important to take into account that it's an unsupervised learning method and that it doesn't require labels. Clustering methods are divided into hierarchical, partitional and bayesian. The first one finds succesive clusters using previously established clusters and partitional algorithms generally determine clusters at once.

K means is a partitional algorithm that assumes that the number of clusters is known. The main idea of the algorithm is to represent clusters with their centroids and find the centroids and clusters that minimize squared distances between elements and centroids. Lloyd's algorithm is an approximation that consists of: choosing k random centroids, assigning elements to clusters using a Voronoi diagram, computing new centroids and iterating until convergence is reached. The convergence is defined when there's no reassignments of elements to other clusters, no variance of centroids or when there's a minimum decrease in the sum of squared error [3].

Other algorithms of partitional clustering are model based like mixture of gaussians (gmm). The main idea of this method is to represent groups with gaussian distributions. In other words, clusters are formed by describing the probability density function of variables by a mixture of multivariate normal densities. Therefore, the objective is to find the parameters of a mixture of Gaussians that explain best the data [3].

Hierarchical methods can be divisive or agglomerative. This methods use dendrograms but the first one starts spliting all the data while the second one starts comparing pairwise data. This paper uses agglomerative clustering which starts merging the most similar pairs of clusters and stops when there's only one single cluster conformed by all the data. Watershed is a grouping method for gray level images that considers the graphs of images as topographic surfaces. The domain of the functions are divided into catchment basins of regional minima and these are divided by watershed lines. The main idea is that a regional minimum can be represented as a whole in the topographic surface. Then the surface is flooded and when two lakes of different holes merge, a watershed line is constructed to separate the regions. However, images are generally over segmented so it's recommended to use marker based watersheds by imposing minima in the marker's positions. A way of obtaining automatic markers is by specifying a height h were regional minima are connected by a path of this height [3].

The objective of this paper is to create a region based segmentation function that uses as features the color spaces rgb, lab and hsv and also x and y spatial components if desired. Also, the method of segmentation can be chosen from k means, watershed, gmm and hierarchical. Aditionally, an evaluation method was proposed.

2. Materials and methods

The implementation of the segmentation method is divided in two main parts, the feature extraction in each colorspace and the segmentation process by each clustering method.

Feature space

The method has six possible feature spaces. Three for the channels of each colorspace, rgb, lab and hsv and the other three for the channels of each colorspace plus the spatial x and y coordinates of the pixel. First, to obtain the image in lab and hsv representation we used the Matlab built-in functions *rgb2lab* and *rgb2hsv*. Then, we iterated through every pixel of the matrix to extract the features, whether it were only the three channels of the colorspace or the 5 channels including location. For the spatial coordinates we assumed that the upper left corner of the image was the pixel (1,1) and went through the whole image until the last pixel ath the bottom right corner. With these features we created a vector of m*n rows (m being the number of rows of the original image and n of columns) and the number of columns of the feature vector.
We normalized the values of the feature vectors in order to get a feature space with a coherent magnitude. For example, in Lab colorspace the value range for each channel differ, L goes in a range of [0,100] while a and b go in a range of [-128+127] [2]. To modify the magnitude of this space we added 128 to the value in a and b. Then we divided by 100 for the L channel and by 255 for the other channels. By normalizing the featurespace we ensured that all channels have the weight in the segmentation.However, we found that the features that included colorspace and spatial position needed a downscaling in the spatial coordinates. This was done by dividing the value of the parameters by a factor of 2.

However, since watershed clustering method uses a matrix of the size of the image as parameter we decided to make a matrix out of the average values of the feature vector for each pixel. This matrix will not give results with the same accuracy as the features vector because it does not take into consideration the fact that similar matrix values might be due to very different values on the same channel but similar values on different channels.

Clustering

For the segmentation phase, four different clustering methods were used, kmeans, gmm, hierarchichal and watersheds. The number of clusters for the output segmentation were determined by a cualitative analysis of each image in the dataset.

For kmeans, we ran the features matrix using the matlab function *kmeans* with 5 replicates to ensure that the convergence of the method was accurate.

gmm was done by first creating a gaussian mixture model with the data in the features matrix via *fitgmdist*. Then, we used the function *cluster* to find the cluster configuration that would best adjust the data to the gaussian mixtures. Finally we obtained a vector of labels and the reshaped it with the same size of the original image.

For watershed, the gradient was obtained using a sobel edge mask and the gradient magnitude was computed. Then, h minima were used as markers for watershed, the amount of labels of clusters was obtained for the specific h value and the process was repeated until the number of minima were equivalent to the amount of desired clusters. This method didn't work well for images of the database because the edges obtained with the gradient weren't closed. Therefore, the h value was qualitatively determined instead of iterating to force a number of clusters in the watershed method.

First, for the hierarchical method it was necessary to rescale the images due to lack of memory and high computing time. Then, an agglomerative tree was constructed implementing the *linkage* function. Then, clusters were obtained employing the *cluster* function and the level of the tree used for the segmented image was determined by the number of clusters parameter.

Evaluation

For the performance evaluation of our method we decided to implement our own strategy. In this strategy we first have one of the groundtruth segmentations, we choose the first labeled region of the segmentation and use it as a mask to extract the same region in the segmented image created by our method. Then we calculate the number of pixels in that region of the segmented image that have the same label and we make a ratio of # pixels of the label in Ground truth over # of pixels with the same label in that section of segmented image. We calculate the ratio for every label of the groundtruth and the make an average of the points obtained for the image. Finally, to incorporate all versions of the groundtruth we run the same evaluation with each of the segmented ground truths and the calculate an average.

3. Results

Because of the varying types of images in the dataset we decided to analyze the results of the methods with only one image from the dataset. By doing this, we can ensure that the difference in outcome of the method is due to the parameters of featurespace and clustering method and not because of the differences between images. We chose the image of the church because it did not have pronounced textures (since we are not using a texture descriptor) and it was possible to qualitatively select a number of 3 clusters.

3.1. K means

In figures 10, 11 and 12 we can observe the differences between segmentations using rgb, Lab and hsv color spaces respectively. The overall results of these three methods are acceptable, each method is able to segment the church from the background and then segment the stairs from the rest of the church. However, it is possible to observe that the hsv segmentation recovers more details than the other two methods. This results show that, for this certain image, an hsv representation with each of the channels as features is the best possible configuration for kmeans.

Additionally, in figures 1, 2 and 3 it is possible to see the results of the segmentations for each colorspace including the spatial coordinates in the features vectors. To illustrate the importance of scaling the values of the channels we scaled down the spatial channels of the rgb space but left unscaled the same channels in the other colorspaces. The results for rgb with spatial coordinates are better than the original method, this was achieved by giving a smaller weight in the feature vector to the x and y coordinates and a bigger weight to the rgb channels. On the other hand, the results for Lab and hsv are not accurate. With this outcomes we can see that the scaling of the spatial coordinates is a crucial factor in obtaining correct a correct segmentation. Downscaling is important because the spatial coordinates are used to support the similarities between pixels, pixels that are next to each other are more likely to belong to the same category, but this does not give as much information of the image as the channel values. Therefore, to achieve the best results with kmeans we have to use the colorspace that allows to differentiate between segments, in this case hsv, and then use the spatial coordinates as support to the classification.

Also, it is possible to identify that the evaluation method is not precise, since the worst result 3 gives the higher value.

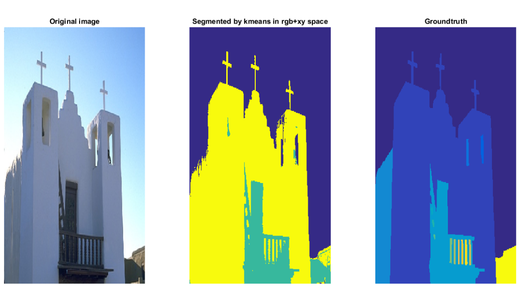

Figure 1. Image segmentation for kmeans with rgb+xy colorspace. Evaluation metric=0.4469

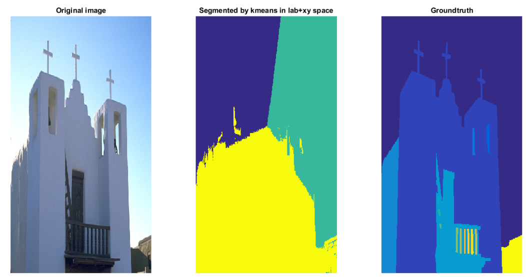

Figure 2. Image segmentation for kmeans with lab+xy colorspace. Evaluation metric=0.4469

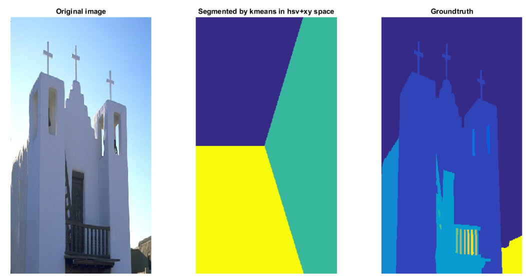

Figure 3. Image segmentation for kmeans with hsv+xy colorspace. Evaluation metric=0.8

3.2. GMM

Based on the outcome obtained with this method, we are able to identify that, once we have made the weighting of the spatial features, the results given by the vectors that include spatial information are better than the ones that do not include it. In figure 15 the segmentation method was not able to extract the church from the background. However, when spatial information was added to this vector it was possible to obtain not only the church but also other details that were lost in the first attempt.

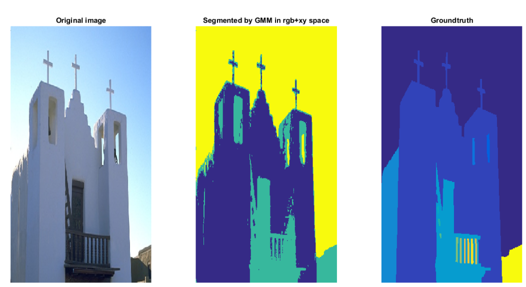

Figure 4. Image segmentation for gmm with rgb+xy colorspace. Evaluation metric=0.44776

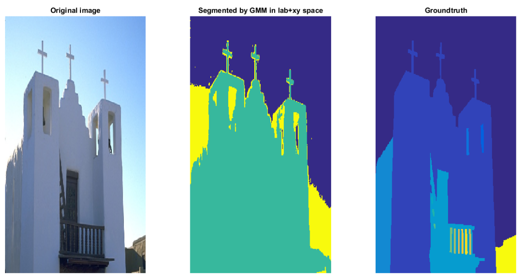

Figure 5. Image segmentation for gmm with lab+xy colorspace. Evaluation metric=0.6532

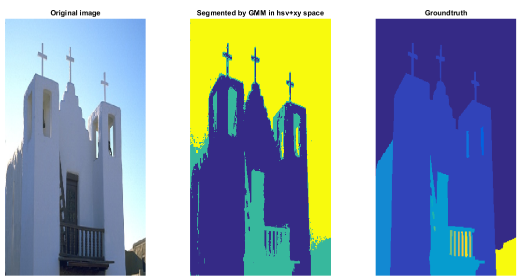

Figure 6. Image segmentation for gmm with hsv+xy colorspace. Evaluation metric=0.5207

3.3. Watershed

The most significant limitation observed for this method was obtaining the regional minima markers required to impose them to watershed. It's important to take into account that imposing minima is of extreme importance when implementing watersheds to reduce the superpixels output. However, markers are difficult to select because they're problem specific. However, for this paper, the h minima were selected as markers to impose minima and also to control the amount of clusters generated. The problem with this method was that when calculating the h minima transform, Matlab defines regional minima as connected components of pixels with constant intensity value and that have external boundary pixels with a greater intensity value. So, as it can be observed in figure 7, when calculating the gradient of the image, a lot of the edges aren't connected. This results in good segmentation in small regions that aren't as important as bigger structures but that have connected edges such as the windows in the church. On the other hand, it's possible to observe that the image of the coins has well delimited edges and therefore has an excellent watershed segmentation using the h transform.

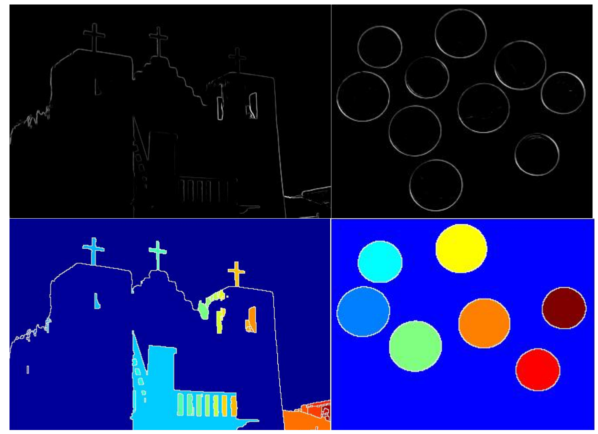

Figure 7. Comparison of watershed segmentation of two images based on their edges in the gradient.

It's important to take into account that the segmentation of the church image in figure 7 was done by manually selecting an h that yielded good results. For automatic h value detection to impose the amount of clusters for watersheds, the results were completely unacceptable (figure 8). This is due to the problem of region segmentation of closed edge regions. For future studies it's recommended to dilate borders in order to close edges. However, choosing the structuring element might be difficult in order to be used for different types of images.

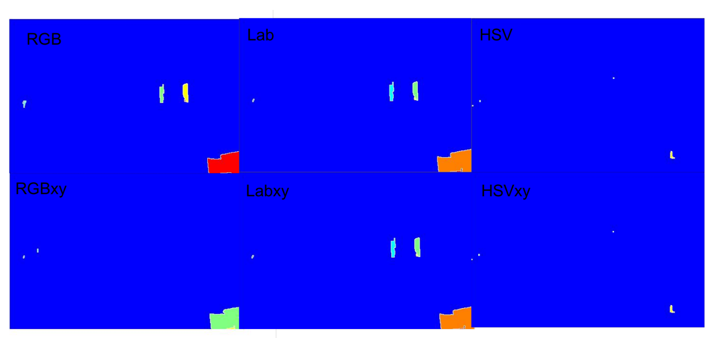

Figure 8. Watershed results for the feature spaces RGB, Lab, HSV, RGBxy, Labxy and HSVxy using h automatic calculation. The evaluation values were respectively:0.8805, 0.8807, 0.9993, 0.9338, 0.8807,0.9993

Figure 8 illustrates that there were no significant differences between feature spaces including only color features and the ones that considered spatial position of the pixels. This results weren't expected because the spatial information gives additional information necessary to distinguish between near pixels with different colors or viceversa. Nonetheless, the weight of the spatial coordinates was only 20% while the colors were 80%. Due to the fact that for watersheds a 2 dimensional image is required, the features were averaged instead of taking into account the entire vector of features. This might have significantly affected because color information is being mixed with spatial data.

Additionally, based on the evaluation results presented in figure 8, we can see that the evaluation method proposed isn't very appropiate for the cases were the majority of the image is labeled with the same value.

3.4. Hierarchical

An important limitation of this method was that it was very slow for large scale images. Therefore it was necessary to resize the images. Howevere, as it can be observed in figure 9, important information of the borders and object's shape is lost. Also, large amount of clusters require to resize even more the image which results in loosing a lot of information. However, it's easy to change the maximum number of clusters desired because the tree is already constructed. To obtain more accurate results it's recommended to study other methods that describe features in smaller size vectors.

Figure 9. Hierarchical segmentation with the image resized to 50% and 2 clusters used. The evaluation metric yielded 0.8853.

4. Conclusions

It is possible to identify that the evaluation method is not coherent with the qualitative results. This missmatch in results and metric is because it only takes into account the pixels of the segmentation that are the same in a certain region of the groundruth but it does not analyze the pixels of the groundtruth that are the same in certain label of the segmentation. Our evaluation method works well for various categories in which there will always be a small region corresponding to the cluster. However, when the image is segmented in only 2 parts, the regions are big and the metric fails to give a correct metric. This problem can be fixed by adding to the metric the pixels of the groundtruth that are the same in certain label of the segmentation. This change would ensure that segmentation errors that leave regions of big size would no longer be acceptable.

It was possible to observe the importance of using different color spaces because depending on the application, some channels were more discriminative and this aided for a better segmentation. For future studies it's recommended to try methods with automated definition of optimal amount of clusters. Also, an evaluation method such as precision recall curves could be implemented. The methods should be tested with more images and other features such as texture could be taken into account to obtain better segmentation results.

Finally, it is not possible to give a full comparison between clustering methods since the metric we developed is inconsistent. Therefore, the only way we could give an approximation of which method is better is by qualitative analysis and in segmentation methods a qualitative analysis is not a correct approach.

References

[1] Visual computation and multimedia. *Universidade de Beira Interior*.

[2] Introduction to colour spaces, 2018.

[3] P. Arbelaez. Lecture 5: Clustering. *Universidad de los Andes*, 2018.

[4] D. Naik and P. Shah. A review on image segmentation clustering algorithms. *International Journal of Computer Science and Information Technologies*, 5(3):3289–3293, 2014.

Additional attachments

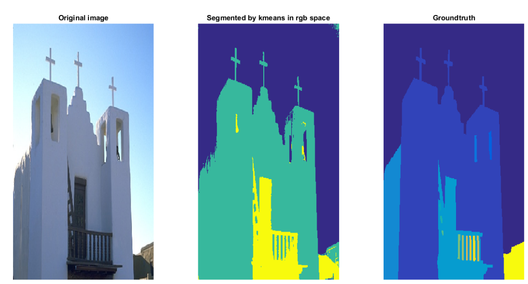

Figure 10. Image segmentation for kmeans with rgb colorspace. Evaluation metric=0.4438

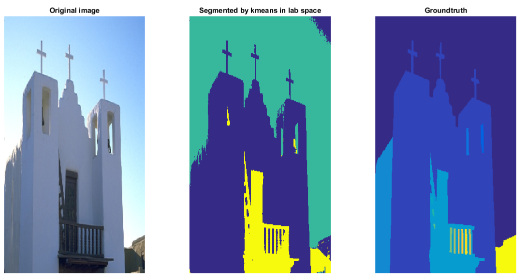

Figure 11. Image segmentation for kmeans with lab colorspace. Evaluation metric=0.6310

Figure 12. Image segmentation for kmeans with hsv colorspace. Evaluation metric=0.6245

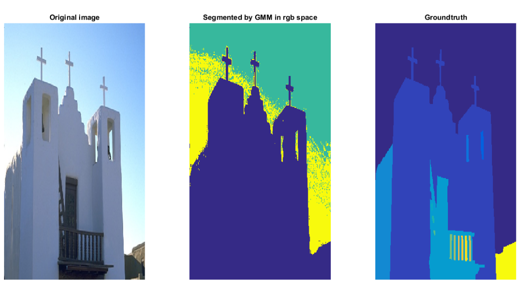

Figure 13. Image segmentation for gmm with rgb colorspace. Evaluation metric=0.8669

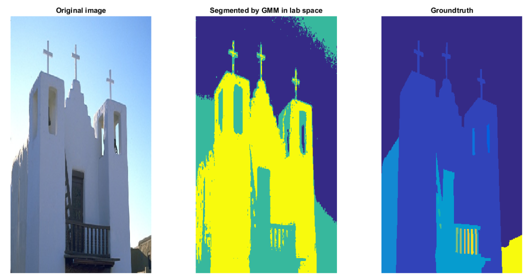

Figure 14. Image segmentation for gmm with lab colorspace. Evaluation metric=0.7108

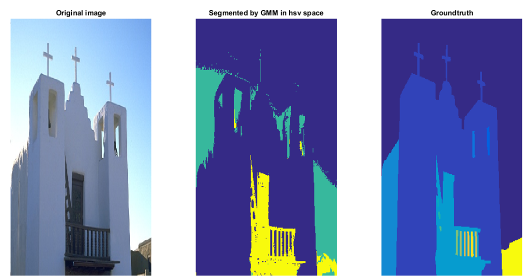

Figure 15. Image segmentation for gmm with hsv colorspace. Evaluation metric=0.6245