Unsupervised Clustering segmentation by K-means, Gaussian mixtures, Hierarchical grouping and watersheds using color and spacial features in a subset of BSDS500

Stephannie Jimenez Gacha
Universidad de los Andes
Cra 1 Este No 19A - 40 Bogotá, Colombia
s.jimenez16@uniandes.edu.co

Sergio Galindo León
Universidad de los Andes
Cra 1 Este No 19A - 40 Bogotá, Colombia
sa.galindo10@uniandes.edu.co

Abstract

Using a subset of 24 images of the Berkeley Segmentation Dataset and benchmark 500, four different unsupervised methods of clustering were developed. Those techniques were made up of K-means, Gaussian mixtures, hierarchical grouping and watersheds. The color information and spacial features were as well part of the segmentation. Afterwards, the average Jaccard Index was proposed for evaluating the segmentations made from each one of the results. The best method using this metric, was the Gaussian mixtures using both color and spatial information with an index between 0.33 and 0.43. The hierarchical and k-means offered comparable results. The method for evaluation could be improved by comparing the segmentation with the annotation that has the highest affinity.

1. Introduction

Segmentation can be seen as a refinement of classification problems where the main objective is to classify each pixel of the image into a group belonging to a semantic category. In an image, multiple categories or objects can be found, however, at least two different categories should be present, background and object, to perform a segmentation.

As segmentation corresponds to a pixel classification inside an image, multiple classifiers and feature spaces can be useful. In general, objects in images can be described as a collection of pixel with similar color, texture and space location. For that, color spaces and spacial information have been used as features to construct a representation space, like *lab+xy*, that combines both types of information[1].

After that, to classify each pixel in the feature space, unsupervised clustering algorithms as k-Means or hierarchical groping are used[2]. Even though supervised algorithms can be used for segmentation purposes, if annotations are inexistent or insufficient or if time corresponds to a significant factor, the usage of unsupervised algorithms is preferred. Also, unsupervised classification might be the first approach to a unknown image whose objects wanted to be retrieved[2].

For that, an implementation of a segmentation method by means of unsupervised classification using different clustering techniques and representation space is proposed. The method will evaluate the performance of k Means, Gaussian mixtures, hierarchical grouping and watersheds combined with 6 different feature space, 3 corresponding to pure color representation and 3 corresponding to color+spacial representation over a subset of 24 images of the BSDS500 dataset.

2. Materials and Methods

To create and evaluate a segmentation method using different feature spaces and clustering algorithms an executable file was developed. The file contains two functions, one to load an image from the dataset as well as the main segmentation function. The file also contains the evaluation process that was performed for the entire set of images varying the feature space as well as the clustering method.

The subset of the BSDS500 used corresponds to 24 random images selected from the main BSDS500 full dataset with their three corresponding human segmentations. The images have a size of 481x321 pixels and are in .jpg format, like all BSDS images.The images re found in landscape and portrait orientation.

2.1. loading and Segmentation function

In order to create the segmentation function the image loading function was created first. This uses the PIL and numpy libraries to load all images as uint8 arrays. After that, the main segmentation function was constructed.

At first, the correct color representation was done using the color package from the skimage library according to the specified color space given as a user-defined parameter. Regarding the lab and hsv color spaces, as the default incoming image is assumed to be RGB (RGB:[0-255]) and they map to the [0-1] range, all three representations were normalized to the a specified value to avoid a greater significance of one of the channels. If the representation space corresponds to a color-spacial representation, the x coordinate and y coordinate of each pixel was added to the representation space. Also, to avoid a spacial greater significance, this two additional channels were normalized into the same range as the color spaces.

Then, the clustering method was selected by means of another user-defined parameter. The clustering methods implemented correspond to K-Means, Gaussian Mixture, hierarchical clustering and waterseds. The K-Means method calculates the location of k centroids in the feature space by reducing the distance to each point in the space to its' corresponding centroid such that, in each iteration, new centroids and labels are assigned based on a distance metric provided. This method tends to return circular and homogeneous clusters. The K-Means clustering method was implemented using a user-defined number of clusters and a deterministic state.

On the other hand, the mixture of Gaussian clustering method, that assigns probabilistic labels using parametric distributions to fit the cluster shapes and that, as K-means, uses a distance metric to minimize the distance between each point and its' centroid was implemented using a full covariance matrix and user-defined number of clusters. Also, the hierarchical clustering method, that joins or separates pairs of points in the feature space based on minimum given distance metric was implemented as an agglomerative method using the euclidean distance.

Finally, the watershed method, that finds the border regions when a flooding in the image, seen as a topographical surface, occurs, was implemented by finding the same number of regional minima as number of clusters defined by parameter. The markers of the method corresponds to n-local minima, that were found using the peaklocalmax() function of the skimage package over the reflection of the image respecting the third dimension after averaging in the third dimension all the channels.

To get the segmentation, the nxm images were reshaped into nxm points of the feature space and given to the desired clustering method to obtain the label of each point. Then, the indexed image was compared to the annotations provided.

2.2. Evaluation Methodology

As the hierarchical agglomerative clustering needs a high computational power and memory as well as a great quantity of time, the images were downsampled to 100x100 to perform the segmentation by agglomerative clustering and upsampled to the original size for evaluation.

One of the most important parameter selected was the number of clusters that was fixed to 5 or 10. This values were selected because in average, there are 5 principal objects in each scene and the annotations have in between 20 and 25 different objects per scene. As much of the annotations corresponds to small and insignificant objects the number of clusters was selected 5 because of the number of principal objects or 10 as the half of the average number of groups per annotation.

Finally, for each image, the segmentation using each clustering method and feature space was obtained and evaluated against each annotation. To obtain a proper evaluation methodology, the Jaccard index between each human annotation and each computer annotation was obtained. Each region was compared to all others and the best Jaccard index for each region was stored. Finally, the Jaccard index of a single image was obtained by averaging the best k (k=number of clusters in the image) Jaccards and the general index of the method was obtained as the mean of the 24 Jaccards of the whole dataset. The regions' borders were not considered in the metrics.

3. Results

3.1. Image segmentation

The segmentation of the images using different feature space and clustering methodology was performed. The results for different feature space for the same image are shown in the following images. In first place, the image 6 and 3 were obtained using k-means as segmentation method. The main difference, is that the first image was obtained taking into account the spatial information. It can be seen that spacial information improves the segmentation as it incorporates one of the characteristic of a visual object, similar spacial coordinates.

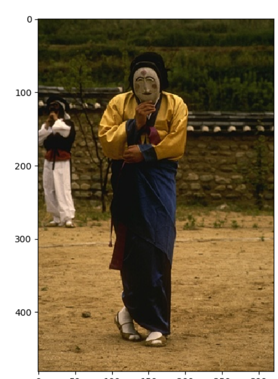

Figure 1. Original 55075.jp image.

On the other hand, the usage of different feature space allows to have different segmentations as the feature space

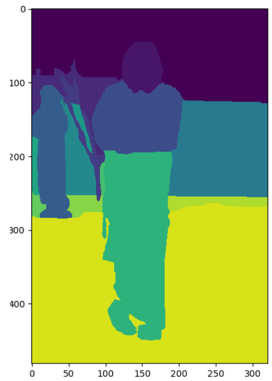

Figure 2. Groundtruth of the 55075.jpg image.

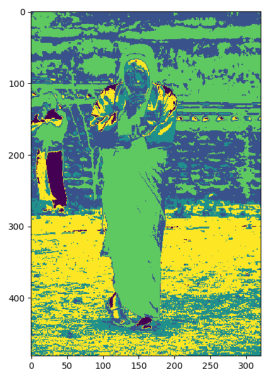

Figure 3. Segmentation of '55075' image using RGB space and K-means clustering (k=5).

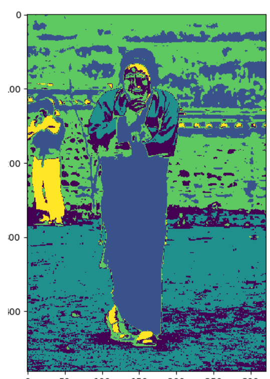

Figure 4. Segmentation of '55075' image using lab space and K-means clustering (k=5).

is representing different characteristics. As objects posses similar colors the usage of lab or hsv color space might result more discriminative while rgb will perform adequate for pure red, green or blue components.

Also, the results for different clustering methods are shown:

From the segmentations, it can be inferred that the k Means, GMM and hierarchical clustering behave adequately and similar as they present more clearly the objects'

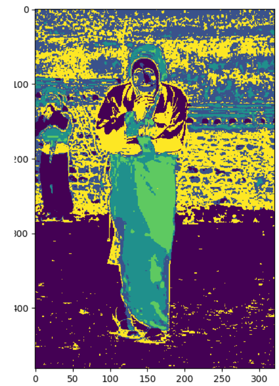

Figure 5. Segmentation of '55075' image using hsv space and K-means clustering (k=5).

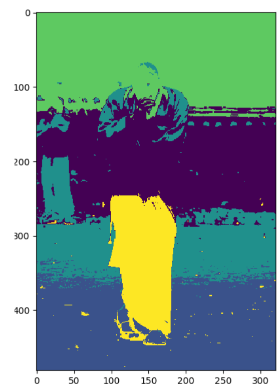

Figure 6. Segmentation of '55075' image using rgb+xy space and K-means clustering (k=5).

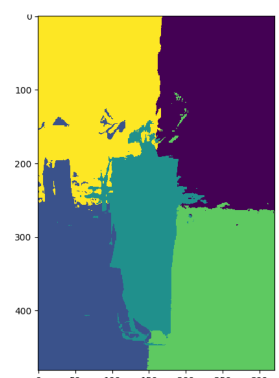

Figure 7. Segmentation of '55075' image using lab+xy space and K-means clustering (k=5).

regions. The watershed method could have some problems in the regional minima selected, and also, the averaging in the third dimension to obtain a topographical surface might not be an appropriate option as some fundamental object information can be lost or modified. Also, for watershed segmentation the channels representing the grayscale or luminosity information of the images seems more discriminative than the remaining channels (like the L channel in the

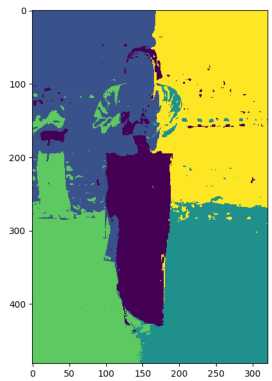

Figure 8. Segmentation of '55075' image using hsvxy space and K-means clustering (k=5).

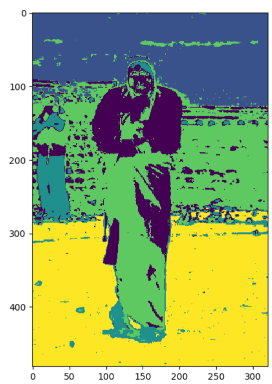

Figure 9. Segmentation of '55075' image using rgb+xy space and GMM clustering (k=5).

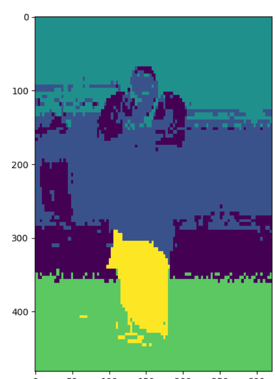

Figure 10. Segmentation of '55075' image using rgb+xy space and hierarchical agglomerative clustering (k=5).

lab color space), also, the spacial information adds noise to the minima selection.

3.2. Segmentation evaluation

The Jaccard average index for each feature Space and method was obtained using 5 and 10 clusters with spacial and color features normalized to the same value. Also, the Jaccard index for each feature space and method was

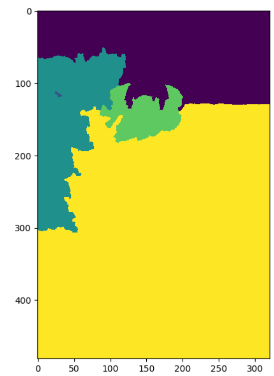

Figure 11. Segmentation of '55075' image using rgb+xy space and watersheds(Imposed minimum =5).

obtained using 5 clusters and a 5-fold space normalization value respecting the color feature normalization value. The results are shown. To compute the Jaccard index the Python3 logical AND and OR were used to obtain binarized masks of each region intersections and unions between the human and computer segmentations. We used the formula $J = Intersection/Union$. The borders were not considered in the evaluation metrics.

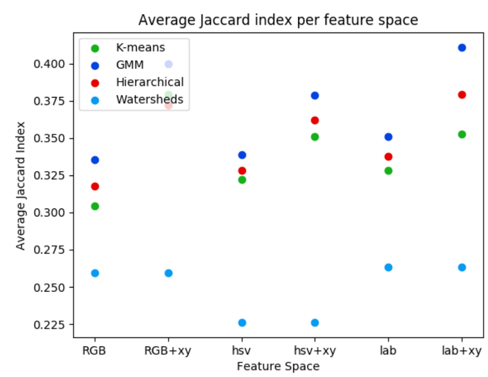

Figure 12. Jaccard index by method and feature space with 5 clusters and equal color and spacial normalization value.

The general evaluation strategy for a segmentation problem corresponds to the Jaccard index over the complete set of annotations, however, the proper evaluation for multiple ground truth corresponds to adaptive ground truth composition were the segmentation is made up of different parts of the ground truths and the final Jaccard index is calculated averaging the area of each mask used[3].

For that, the implemented evaluation strategy might work properly it the segmentation produces medium size compact objects. If the segmentation produces big spreaded objects over the scene the Jaccard index might be above reality as it intersects with all annotations. The selected evaluation strategy is in accordance the general segmentation evaluation as it uses the Jaccard index, also, it allows each

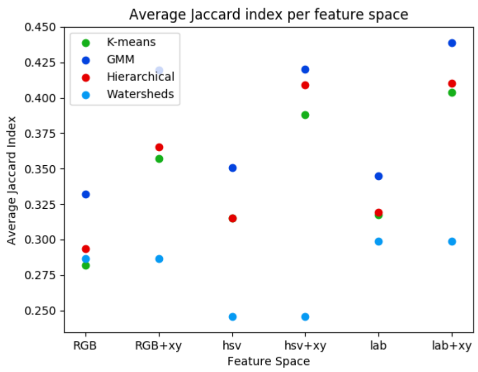

Figure 13. Jaccard index by method and feature space with 10 clusters and equal color and spacial normalization value.

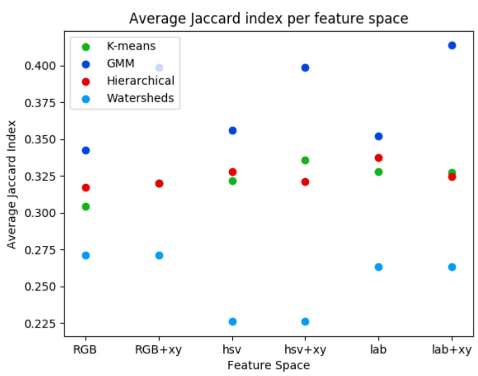

Figure 14. Jaccard index by method and feature space with 5 clusters and spacial normalization value = 5x color normalization value.

segmentation to contribute equally and to be contrasted respecting all annotations, which is fair, when unsupervised clustering is performed.

Using our evaluation, it can be seen that the GMM corresponds to the best clustering method with an average Jaccard index between 0.33 and 0.44 in all feature spaces and above all other methods following by hierarchical clustering and k means with similar results. Also,figures 121314 evidence that the number of clusters directly affects the performance as well as the incorporation of spacial features. An increased number of clusters (10), closer to the actual number in the groundtruth (20),increased the performance 0.05 in average. Also, it is important to notice that the feature space containing spacial information performs if not as well, better than the color feature space alone.

According to our evaluation,the best results were found using the lab+xy feature space with the GMM clustering method.

4. Discussion

A correct segmentation using unsupervised clustering was performed using the color and color+spacial feature space as they correspond to two of the main characteristics of the visual objects: similar color and similar spacial location. It can be seen that lab and hsv color spaces are more discriminative as they incorporate a grayscale representation of the image as well as color information. Also, the incorporation of the spacial information improves the performance of the segmentation as color+location forms an adequate descriptor for an object in a scene.

The best results were found for GMM using lab+xy feature space. This can be explained to the distance metrics used by the GMM method that offers more naturally grouped clusters than the other methodology and because of the color space used. Watersheds method presented inconveniences in its' implementation as the average along the third dimension of the 3 or 5 channel descriptor is not the best option to generate a topographical surface. A weighted averaging is suggested.

One of the weaknesses of the method rely on the dependence of the segmentation on the color information. An image with definite regions using a restricted area of the color spectra will be inadequately segmented. Also, the method will perform better on rounded or compact objects whose pixels are close to the center of mass of the object. Animal legs or object prolongations might be segmented in an improper way.

Finally, the evaluation strategy can be improved by comparing each segmentation obtained to the more alike annotation mask. Also, the evaluation should consider the number of actual objects in the image and the number of objects obtained in the segmentation to weight the contribution to the Jaccard index. An index of 0,5 with 2 segmented objects should be different to the same value obtained with 10 objects. Also, the results can be improved by increasing the number of clusters to a closer value of the actual number in the annotations and by including texture in the feature space as it corresponds to another important characteristic of visual objects.

5. Conclusions

An unsupervised clustering algorithm using a feature space composed of the color and spacial information of the pixel can be used to segment objects in images. This method is based on the characteristics of visual and spacial homogeneity of objects as visual entities, that tend to share color and spacial information in a scene.

The best results are obtained when using a Gaussian mixture clustering method with the lab+xy feature space. This is explained by the joint of both color and spacial characteristics of pixels using a discriminative color space, with a more flexible clustering algorithm. In addition, the hierarchical and k means clustering offer comparable results respecting the GMM, joining similar pixels in a more rigid and predetermined way.

It is possible that the dimension reduction to process the images in hierarchical clustering and the final up-sampling add noise to the segmentation as the borders become less defined due the nearest interpolation used. Finally, the watershed clustering method shows poor results due to some characteristics of its implementation like the regional minima selection.

The evaluation proposed uses the Jaccard index following the segmentation general evaluation, however, it can be improved to avoid comparing each segmentation and annotation but comparing each segmentation with the corresponding annotation. Also, the averaging function should be weighted to consider the number of objects and the multiple annotations.

The color and spacial channels scale must be preserved in a moderate ratio of 1:5 as maximum to obtain adequate results. If the color channels are scaled up significantly, the spacial information is neglected while, in the opposite case, the color information is neglected leading to a uniform partition of space. Finally, the method can be improved by increasing the number of clusters and by including the texture information, that corresponds to an object related feature.

References

[1] G. Coleman and H. Andrews. Image segmentation by clustering. *Proceedings of the IEEE*, 67(5):773–785, 1979.

[2] 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, 1 2015.

[3] B. Peng and L. Zhang. Evaluation of Image Segmentation Quality by Adaptive Ground Truth Composition. pages 287–300. Springer, Berlin, Heidelberg, 2012.

6. Code and Images

The code is available at the team's repository at *https://github.com/steff456/IBIO4680/tree/master/06-Segmentation/Answers*