Segmentation

Jaime Enrique Cascante Vega
Universidad de los Andes
je.cascante10@uniandes.edu.co

Angela Castillo Aguirre
Universidad de los Andes
a.castillo13@uniandes.edu.co

Abstract

The segmentation of objects within an image is an open challenge that has been studies among the last years. Some of the methods used for this purpose is the clustering. For this study it was considered the k-means, Gaussian mixture models, agglomerative hierarchical and watershed methods to do the segmentation of objects within an image. It was also taken into account the RGB, Lab and HSV space colour and, in some cases, the spatial location of the pixels in the image. The results showed that the clustering can actually work as a segmentation method, but its performance will depend on the hyperparametrization of the functions in order to achieve good results. However, assessed with a proposed evaluation metric, the method whose segmentations have the least difference with the ground-truth is watershed under the colour space RGB considering the spatial information of the pixels.

1. Introduction

Automatic object segmentation in images is one of the actual challenges in computer vision. The geometry perception grouping and the perceptual organization of an image, is one of the tasks that researchers have been study the last few decades. [1] The human perception allows the observer to identify different regions on an object, even if there are not clear complete edges of it, because the brain is in charge of the visual function and is able to interpret an image from the information from the environment. [2] Because of this, there are some strategies implemented in order to identify objects in images using unsupervised segmentation algorithms performed by clustering methods. On this study, some segmentation algorithms were tested to determine the performance of the segmentation comparing different clustering methods. Techniques such as k-means, watershed, hierarchical and Gaussian mixture model were used to segment images from the BSDS database. [3]

Clustering methods

1.0.1 K-means

K-means method uses the Lloyd's algorithm in which the data is partitioned assigning n observations or points to an specific k cluster. Each cluster is defined by its centroid, which position is changed in every iteration by the average of the observations. At the end, there are defined clusters that characterize a different region in the image. [4]

1.0.2 Gaussian mixture models

In respect to the Gaussian mixture models, the k multivariate normal density distribution components compose this method. The components that describe the procedure consider a d-dimensional mean, a covariance characterized by a d-by-d matrix and a mixing proportion. The method computes different probabilities for the component membership, that contain the relation of probability for each observation. The separation of the clusters is associated to the estimation of the component means, covariance matrices and the mixing proportions. [5]

1.0.3 Agglomerative Hierarchical

On the other hand, the clustering by the agglomerative hierarchical method generates nested clusters by merging them. This hierarchy is represented as a tree, in which the root of the tree is a single cluster that contains the all the observations and the leaves comprise only one sample. The ward method, which is the used for this case, considers the sum of the squared differences to minimize this result for all the clusters. The intuition of this method is based on the k-means procedure, but is implemented using a hierarchy. [6]

1.0.4 Watershed

Finally, the watershed algorithm consists in the division of the image by segments in which it is located a local minimum. The followed idea is that the limits of the regions are the maximum, which at the same time separate a minimum of another. Where one of these maxima is found, a partition is built that does not allow the passage of ”water” from one minimum to another. At the end, there is a complete map of the image with local minimums that represents the segmentation regions. [7]

2. Materials and Methods

2.1. Database description

For the aim of this study, it was provided a reduced version of the BSDS500 database. This dataset contains natural images each with its corresponding annotations done manually by 5 different humans. The annotations have both a ground-truth segmentation, where it is possible to see the different regions of the image, as well as 5 different contours annotations of the boundaries that delimit the different regions in the image. [8]

2.2. Segmentation methods implementation

The images were downloaded from the provided database. Either the images, the segmentations and the boundaries information were stored in variables for further processing. The entire data set available for this study was down-sampled and rescaled to a minor size to achieve several segmentation results. In general, the proposed methods to perform the segmentations require a high machine capacity to be successful. Due to the dataset, having more than one image to process, it was necessary to reduce the total size of the images. Also, the used software for the development of this work was *Python*.

It was developed an algorithm in which a function was created to segment images using different clustering methods. The implementation of the function, first considered the different space colours that can be selected for the analysis of the onset image. For this, the options consisted in the *RGB*, *Lab* and *HSV* space colours. Furthermore, there was an extra parameter taken into account which included the x and y position of each pixel from the onset image. This last parameter gave the spacial information of the pixels to have a better approximation adding this to the clustering phase.

For the k-means method, it was considered a random selection of the initial centroids from the data. Also, the maximum number of iterations per run, was settled at 300.

Regarding the Gaussian mixture method (GMM), the parameters mainly comprised the own general covariance matrix for each component, and a random initialization of either the weights, the means and the precisions. This was made to have a different distribution of the data in comparison to the k-means method.

Besides, for the hierarchical segmentation method it was selected the default parameters of the function which consists on an euclidean affinity as the metric to determine the linkage, and a *ward* linkage criterion to compute the used distance for between the sets of samples.

As well, the watershed method used the negative values of the maximum local peaks, selected with the amount of clusters that enter as a parameter. Former functions were used to implement the final algorithm. [9]

2.3. Segmentation parameter

With respect to the number k of clusters that are used to perform the segmentation, it is considered the number of regions within the image to run the clustering. This indicates that the selected number of clusters for each image varies according to the number of regions that are presented in the ground-truth of that same image. In this way, it would be expected to have a similar object segmentation for each image. Because 5 different segmentations were provided, the strategy to handle multiple ground-truth data was to take one of the annotation images as the ground-truth for the image in which the current segmentation is performed. In this way, the differences between a ground-truth and another are ignored.

2.4. Evaluation methodology

For the evaluation of the clustering methods, it was computed the root mean square (RMS) for both the segmented image and the annotation image. Then, the difference between this measure is determined for each pair of images (image and annotation) of each method combination. That is, according to the colour and the clustering method that is used in the current segmentation, the images will have a value that determines what so close is the RMS of the segmented image against it corresponding ground-truth. In this way, we have a value associated with each image for each clustering strategy. For this, the average of the difference is taken for each combination of colour space and clustering method, and thus a single value is determined for each possible combination. At the end the resulting points are plotted and compared to determine the best result. For the aim of this study, a clustering strategy will be considered good when the value of the difference is smaller, since this indicates to us that the resulting images of the segmentation are more alike to their corresponding segmentation annotations.

3. Results and Discussion

There were 5 different ground-truth segmentations for each image, so it was selected one of the annotations in order to reduce the problem of ambiguity that is given by the number of people who performed different segmentation possibilities for the same image. In addition, the maximum number of clusters in the ground-truth data was chose as the number k of clusters used in the algorithms to segment and, in this way, compare the proposed method with the provided annotations.

The images were clustered using the different possible combinations of the parameters. Some of the results are presented in the following figures. The obtained segmentations are presented in the right side of the image, while the image in the left side corresponds to the ground truth. For instance, figure 1 represents the automatic segmentation using the k-means method in a RGB space colour.

As it is possible to see in figures 3 and 4 the differences between the spaces colours and the actual spatial information of the onset image, can change the performance of the method. This can be seen by the differences in the regions from one image to the other (speaking of the realized segmentations), as seen in figure 3, the pixels marked with the same label, are much more dispersed than the ones are. pixels in figure 4.

In figure 3 the segmentation is partially and contains a lot of errors, while the figure 4 presents the exact same clustering method, varying the space colour and taken into account the spatial information of each pixel.

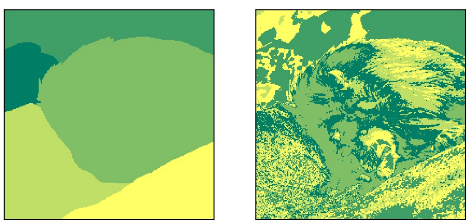

Figure 1: Image segmentation by k-means method using RGB space colour. The parameter k is 6.

Although it is possible to have information about the segmentations through the visual analysis of the images, the evaluation method was applied. It was expected to have some metric that verified or gave more clarity about the segmentations made. With respect to this chosen evaluation method, it is known that the RMS is a measure that calculates the magnitude of a variable quantity, so we this, we pretended to determine the variability among the ground-truth images in order to compare this variability in the resultant segmented image, according to this measure.

Figure 5 (in appendices) presents the total RMS averages by segmentation strategy. Through this graphic, it is possible to see a wider information about the different methods used. As it is possible to see in this graph, there is a tendency, where there is more difference between the RMS of the segmented images against the RMS of the ground-truth

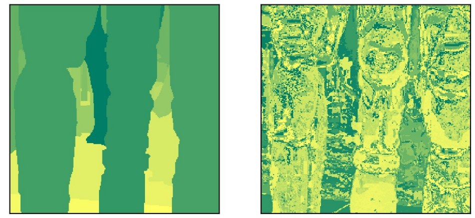

Figure 2: Image segmentation by k-means method using Lab space colour. The parameter k is 11.

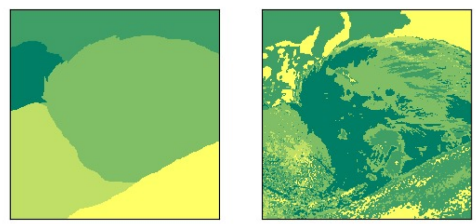

Figure 3: Image segmentation by GMM method using RGB space colour. The parameter k is 6.

images, and this increases when the different clustering and representation methods are varied. In general, it can be said that using this evaluation metric, the clustering method that most differentiates between annotations and segmentations, is the hierarchical. However, as it is possible to analyse, as clustering methods are varied, the measurement indicates that the methods are worse from k-means, passing through GMM to ending up at the worst that is the hierarchical. In addition to having a worsening as the colour space changes, in that order (see the legend of the figure 5), it is possible to see that adding spatial information to the different colour spaces simply worsens the variability within the segmented image, producing a greater difference between annotation and segmentation.

However, although this pattern is very marked in k-means, GMM and hierarchical, this does not happen in watershed, which is very interesting. First, it is possible

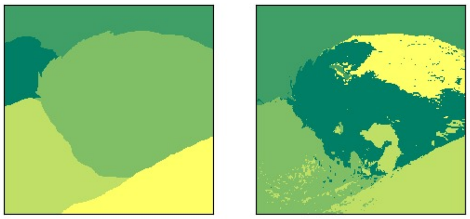

Figure 4: Image segmentation by GMM method using Lab space colour and xy spatial information. The parameter k is 6.

to see that both Lab and RGB do not produce better results than HSV and spaces considering spatial information, which is more noticeable in the other clustering methods. It is also possible to see that the best strategy for this clustering method is to have the RGB colour space considering the spatial information producing approximately 2.8 as the average difference between the ground-truth ans the segmentation.

One of the main limitations of the current methods is the right selection of the hyperparameters because, per each method there are at least 5 different parameters that can me modified and optimized according to the used clustering method. According to the results, the space colour and the information of the location of the pixels inside an object matters. With this, it was proven that the used method is sensible to the position of the pixel either spatially and in the colour space where is described. The best result taking into account the visual analysis of the image, is presented in the figure 4. This may be possible because of the information that the channels bring to the clustering. This is, the channel information plays an important role when the clustering is being performed.

Regarding the evaluation method, it can be said that, although there are measures that could be close to the performance reality of the evaluated methods, in reality it does not consider important information such as the delimitation of the areas that are in the image or the connectivity that is within the segmented regions, as it is considered in the Jaccard index, for example.

4. Conclusions

As it is possible to see from the results, the clustering method to segment objects in images is viable. Nevertheless, the quality of the segmentation will be closely related to the choice of clustering method and the hyperparameterization that is chosen for each of the cases.

Through this study, it was possible to see that the different evaluation ideas could contribute with some information about clustering methods and representation spaces used in this case. This is important because the determination of the quality of a method is subject to this, so the variation of a small part of the evaluation, could cause different results that change the vision on the best method.

References

[1] J. Shi and J. Malik, “Normalized cuts and image segmentation,” *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.

[2] R. P. N. Rao, B. A. Olshausen, and M. S. Lewicki, *Probabilistic models of the brain: perception and neural function*. MIT Press, 2002.

[3] 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,” in *Proc. 8th Int’l Conf. Computer Vision*, vol. 2, pp. 416–423, July 2001.

[4] Matlab, “kmeans.”

[5] Matlab, “cluster.”

[6] Scikit-learn, “2.3. clustering.”

[7] M. Bicego, M. Cristani, A. Fusiello, and V. Murino, “Watershed-based unsupervised clustering,” *Lecture Notes in Computer Science Energy Minimization Methods in Computer Vision and Pattern Recognition*, p. 83–94, 2003.

[8] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, “Contour detection and hierarchical image segmentation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, pp. 898–916, May 2011.

[9] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

Appendices

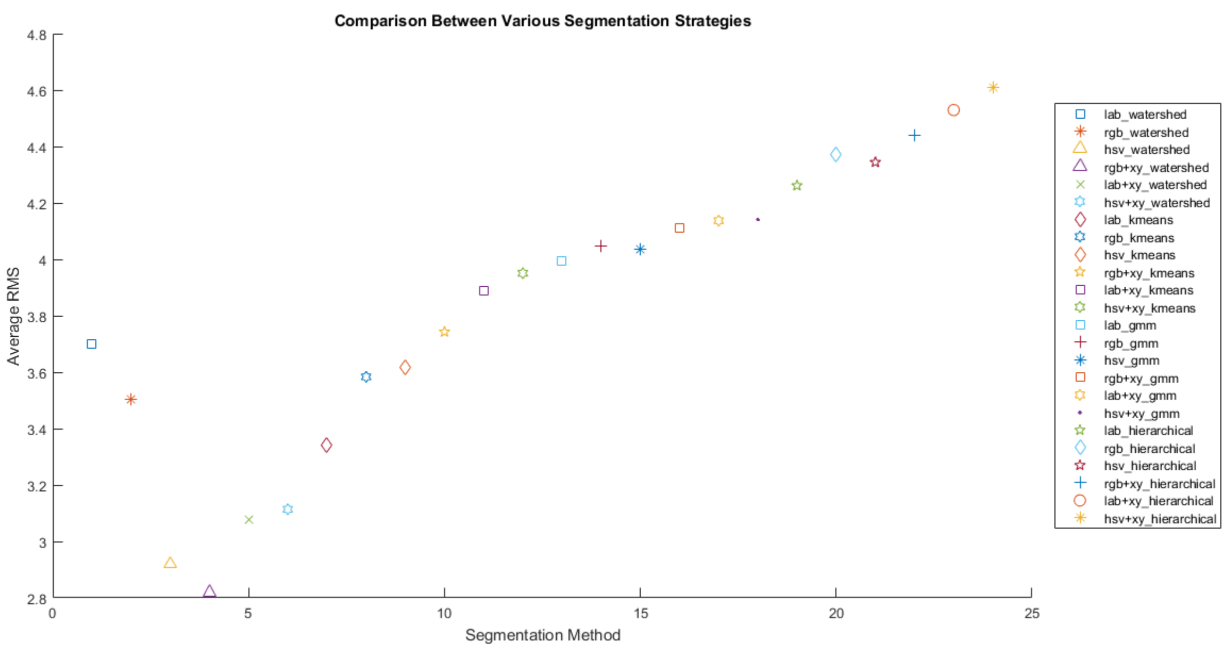

Figure 5: Comparison of the average Root Mean Square (RMS) value between the different feature spaces and clustering methods used to evaluate the methods.