Segmentation

Cristina Isabel González Osorio
ci.gonzalez10@uniandes.edu.co

Raúl Santiago Molina Rodríguez
rs.molina10@uniandes.edu.co

Abstract

In this practice we explored different segmentation methods in samples of the BSD dataset [2], representing the images in multiple color spaces and spatial feature dimensions. A novel segmentation evaluation metric was proposed using local entropy information in the image.

1. Introduction

The ultimate goal of segmentation is to divide the image into regions that belong to objects or entities, ignoring information such as the texture of objects. The problem of image segmentation can be understood as a classification problem in which a label must be assigned to each pixel. From this, it is possible to use different unsupervised classification methods and different representation spaces [1].

2. Materials and methods

2.1. Clustering algorithms

Kmeans
Given a number of clusters K to find, Lloyd's algorithm is used to find the centroids that minimize the variance of points in the same partition (Voronoi cell). It begins with K random means (centroids), assigns each datapoint to a partition according to the Voronoi diagram and then, recalculates K new means, repeating the process all over again. The algorithm iterates until there is no change in point assignment or mean position difference to previous iterations is lesser to a threshold ϵ.

Mixture of gaussians
A more generalized approach of Lloyd's algorithm, it uses Expectation Maximization (EM). Given a parametrized probability distribution (Gaussian), with an initial guess of said parameters, the partial membership (responsibility) to each distribution (K distributions) of each datapoint is estimated (Expectation), then, given the responsibilities, the parameters are estimated again.

Hierarchical
The function *clusterdata* of MatLab implements an agglomerative hierarchical clustering method. In this order of ideas, the clusters with the least distance are merged in a hierarchical way. Starting from the pixels as individual clusters until reaching the number of clusters specified by parameter. In this way, we obtain a family of clusterings represented in a dendrogram which is constructed from individual elements until they are mixed in a single cluster.

Watershed
These method identifies watershed lines from the regional minimums of the image. First, the gradient of the image is calculated. Taking into account that the image has a large number of local minimums it is necessary to obtain a number of regional minimums equivalent to the number of desired clusters. For this, the local minimums of less depth are eliminated with the *imhmin* function.

2.2. Segmentation parameter tuning and Image preprocessing

For the spaces of representation that included the spatial information of pixels, it was necessary to scale down the values corresponding to the positions of the pixels in the image. In the representation space the chromatic information presents a significantly lower variance with respect to the variance of the spatial information, given that a combination of positions xy never repeats. Consequently, if both types of features have the same weight, the classification will be determined to a large degree by spatial information, obtaining clusters that approximate to quadrants in the image, as shown in the figure 1. In Gaussian mixing and hierarchical agglomerative methods, it was necessary to decrease the size of the images since these did not converge to the solution or the memory capacity was overcome. Finally, the number of clusters used for each image was chosen according to the number of objects that were desired to be segmented or the number of labels that were founded in the annotations.

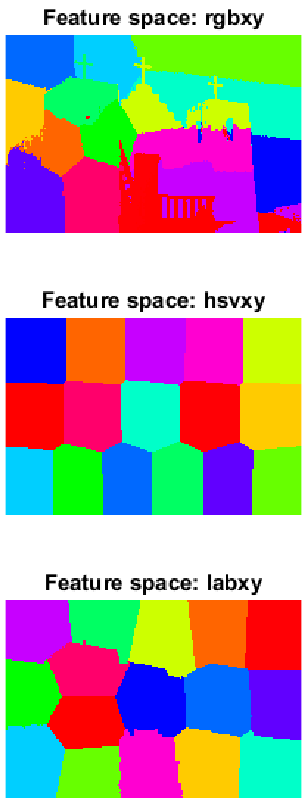

Figure 1

2.3. Evaluation

2.3.1 Framework

To evaluate a segmentation problem it is necessary to compare the proposed regions by the algorithm with the ground truth, however, human segmentation is not a totally consistent task: although most of the salient regions are recognized by various human subjects, these are not exactly equal across subjects. To provide a solid estimate of the correctness of a given method, it is necessary to have multiple ground truths, obtaining an appropriate metric for each one and give a global average (other options might be considered).

2.3.2 Proposed Criteria

For this practice, we developed our own evaluation metric, normalized class entropy difference. The main idea behind this metric is for it to be constrained between 0 and 1, with 1 being total similarity between the prediction and the truth and 0 total dissimilarity.

First, we defined a dissimilarity coefficient c as follows:

$$c(\bar{y}, y) = \frac{abs(H(\bar{y}) - H(y))}{max(H(y), log(K) - H(y))} \quad (1)$$

Where y is the desired discrete distribution and $\bar{y}$ an estimate, $H(x)$ is the Shannon's entropy for any distribution x and $log(K)$ its maximum value for a distribution with K classes (the number of ground truth regions in the image).

Then, a dissimilarity matrix D is computed, such that:

$$D_{i,j} = max(c(\bar{y_{i.}}, y_{i.}), c(\bar{y_{.j}}, y_{.j})) \quad (2)$$

$y_{i.}$,$y_{j.}$ being the discrete distributions of the i_{th} row and j_{th} column of a given image Y.

Finally, a similarity metric s is obtained normalizing the sum of elements in the matrix and subtracting it to 1.

$$s = 1 - \frac{1}{N * M} \sum_{i}^{N} \sum_{j}^{M} D_{ij} \quad (3)$$

2.3.3 Advantages and Shortcomings

An advantage that this metric provides is that it is not necessary to identify the estimated regions with their respective indexes in the ground truth. Although the estimate and the ground truth are compared only once, the entropy computation must be performed $2xNxM$ times, being slower compared to bitwise logical operations as those used to calculate the Jaccard Index. A possible weakness of our metric is the equalization of a random guess estimate. Given that it does not enforce a set number of regions, an image with only one class can obtain a significant increase in its similarity measure compared con a random prediction (Fig. 4).

Another disadvantage is that, to normalize the coefficient c, a reference distribution must be set, which implies that $s(\bar{y}, y) \neq s(y, \bar{y})$. This can be easily fixed using it as an error metric, avoiding the subtraction to 1 and ignoring the normalization term in c, however, for this report, the normalization was kept for an illustrative process.

3. Results

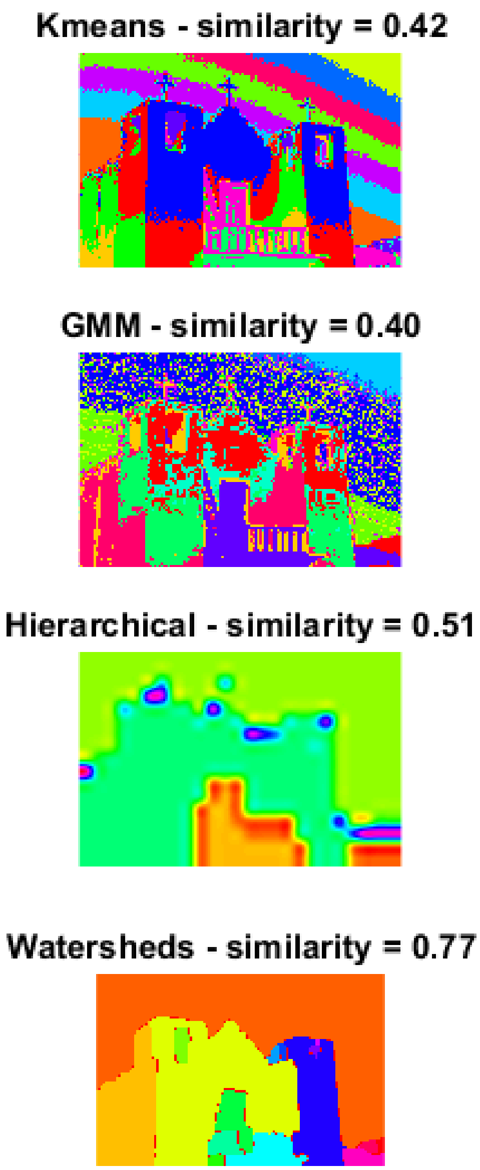

Figure 2

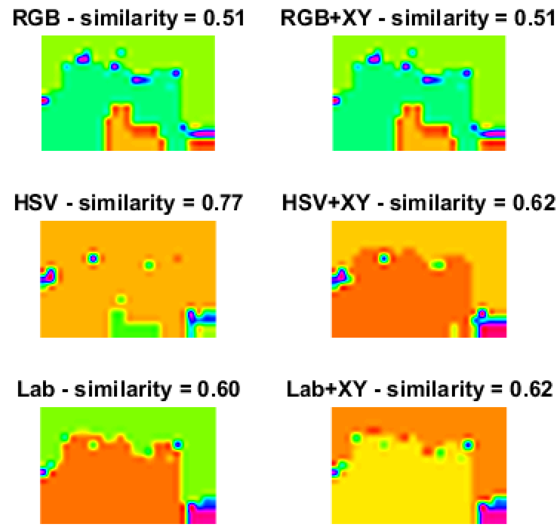

Figure 3

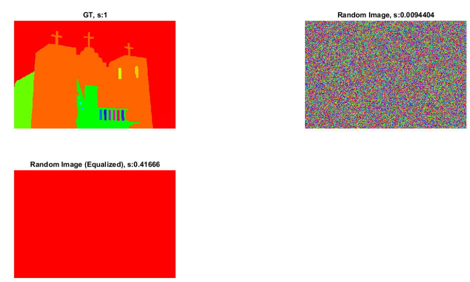

Figure 4

4. Discussion

Due to the colored spaces contain different color information of the image, when comparing the results of the segmentation with different color spaces it is expected to obtain different segmentations, as shown in the figure 3. In this way, the clustering method and the color space to be selected depends inherently on the images and the variety of information contained in them.

With respect to the limitations of the method, on one hand, the results depend significantly on the selected hyperparameters and the information of the image to be segmented. On the other hand, the space of representation could be larger. In this case, the texture information of the objects in the image is ignored. On the other hand, the different methods of clustering used have specific limitations. The K-means and Gaussian mixture methods assume spherical clusters or clusters adjusted to normal parametric distributions, respectively. Regarding the hierarchical method, it temporal complexity is considerably high. Finally, in the watersheds method, the selection of regional minimums may not be done properly. Although the minimums of greater depth are chosen according to the number of clusters, they may not cover all the objects to be segmented or an object may be divided into two or more regions because it contain several minimums.

The greatest fail condition for all methods is the feature space, although some methods might perform better than others, all make assumptions of the feature space. K-means for example, tries to make regions of the same size, producing artifacts in the image that don't make sense (Fig. 2), to a lesser degree the same problem happens for GMM. That problem does not affect watersheds or hierarchical clustering, although both methods might be improved with a better representation of the image. For this it might be easier to use Hierarchical clustering in high-dimensional feature spaces that take into account texture and gradients.

A drawback of our evaluation strategy is that, to normalize the coefficient c, a reference distribution must be set, which implies that $s(\bar{y}, y) \neq s(y, \bar{y})$. This can be easily fixed using it as an error metric, avoiding the subtraction to 1 and ignoring the normalization term in c, however, for this report, the normalization was kept for an illustrative process.

References

[1] R. M. Haralick and L. G. Shapiro. Image segmentation techniques. *Computer vision, graphics, and image processing*, 29(1):100–132, 1985.

[2] 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*, volume 2, pages 416–423, July 2001.