Segmentation in a subset of the Berkeley Segmentation Dataset and Benchmark

Javier Perez
Universidad de los Andes
Bogota, Colombia
je.perezl@uniandes.edu.co

1. Introduction

Image segmentation is a field of computer vision that seeks to extract the coherent entities that are presented in an image [pablo]. This task is based on certain criteria that allows the computer to separate an images in the different objects [pablo].

The Berkeley Segmentation Dataset and Benchmark was created to provide a work frame in image segmentation and boundary detection [cita].

The segmentation problem has been treated as an unsupervised problem as well as a supervised problem. In this laboratory, we performed an unsupervised approach to this problem. In particular, we used k-means, Gaussian mixture models (GMM), agglomerative hierarchical and watershed methods.

2. Materials and Methods

K-means is a clustering method that groups similar data based on a defined similarity. In each cluster formed, all the elements are closer to the centroid's cluster than any other centroid. The number of cluster k has to be specified and the shape of the cluster is round. In an image, each pixel represents a data that has dimension determined by the color space used. For example, a pixel in RGB space has 3 dimensions. This algorithm is done using the Lloyd's.

The Gaussian Mixture Models (GMM) is a generalization of k-means. This algortihm find and asynd different gaussians in order to explain each elements of the data. it produces responslbilies for each clustes, i.e, soft-clusterin [pablo The GMM assumes a normal distribution of the data.

Agglomerative clustering is a type of hierarchical clustering. The clusters are formed by merging the most similar pair of cluster until it reached a single cluster that contains all the elements. This method produces a family of clustering that can be represented in a dendrogram [pablo].

Watershed method is inspired in topographical surfaces and the catchment basins. This algorithms in image segmentation is based on the gradient magnitude of an image that the regional minima should correspond to objects and the closed contours to the watersheds [pablo].

This laboratory was implemented in Python. We defined a function that receives an RGB image, a color space (RGB, Lab, HSV, RGB+xy, Lab+xy, HSV+xy), a clustering method (k-means, GMM, hierarchical, watershed) and the number of clusters (an integer greater than 2). The +xy corresponds to the pixels dimensions in the horizontal and vertical axis. The $(0, 0)$ corresponds to the $(0, 0)$ pixel and the coordinates (x_0, y_0) correspond to the pixel in the row x_0 and column y_0.

We converted the input RGB image to the respective color channel. Because each pixel corresponds to a data, we reshape the images to a matrix NxC where N is the number of pixels in the image and C is the number of dimensions. In LAB, each color channel has different scales, then, we normalize each channel through a bijection to the $[0, 1]$ interval. We also considered the normalization in $[0, 1]$ to each channel when we have 5 dimensions, i.e., RGB+xy, Lab+xy, HSV+xy.

The evaluation methodology consists in comparing each object in the groundtruth to the segmented objects found with the segmentation method. For each groundtruth's object, we obtain the mask of it and we multiply this with the segmented images, where we obtained a non-zero value we have the set of pixels that were correctly segmented. We obtained the fraction of corrected pixels with respect to the total pixels of the mask. We did this for every groundtruth's object, of all the human segmentation labor, and we obtained the average of all the fractions. This number correspond to the accuracy of the segmentation.

3. Results

By observing the images, we noticed that most of them consist of 5-10 objects. There are few cases when the image had more possible segmented objects, and also the humans that performed the annotation had different levels of granularity. We chose $k = 5$ as the number of cluster because it can explain in average the observations described before.

4. Conclusions

References