Unsupervised Image Segmentation based on Clustering and Watersheds

Edgar A. Margffoy-Tuay
Universidad de los Andes
201412566
ea.margffoy10@uniandes.edu.co

Abstract

Uns

1. Introduction

Image segmentation and contour detection are closely related problems that rely on the same premise: Given an image, find a partition (Grouping) of this pixels according to the boundaries that compose it. From this premise, it is possible to treat a segmentation problem as a contour detection and viceversa, this approach allows to propose and define several procedures to solve and evaluate this problem.

Before the introduction of supervised learning techniques to solve this problem, all the approaches and proposed solutions were restricted only to the pixel information present on the image (Unsupervised segmentation) which served as input to different algorithms that provided a higher order representation of the image (Superpixels) or even partial segmentation images. Of these afromentioned algorithms, it is worth to mention clustering-based methods, such as those based on classic generic clustering methods, such as K-Means, Gaussian Mixture Models, Hierarchical clustering (To be explained during the following section) and Mean Shift [3]. Other important methods represent the image as a directed graph with weights in which a minimal cut between regions delimited by a contour is desired, an example of an algorithm present on this category corresponds to the Normalized Cuts framework proposed by Shi and Malik [4]. Finally, other method families try to represent the segmentation problem from a morphological perspective, in which the image is seen as an intensity surface that describes valleys and intensity peaks, in which, boundaries can be seen as inflection peaks, this is the base of the Watershed transform method [2].

Although all the unsupervised methods yield reasonable results that group the pixels of an image into disjoint regions, these methods underperform the performance presented by the modern state-of-the-art supervised segmentation algorithms, which render the previously approach only of interest if a superpixel representation of an image is desired, also, some of the unsupervised models are suitable to semi-automatic applications, such as interactive segmentations.

2. Materials and Methods

About the Implementation

A total of four unsupervised segmentation routines were implemented and subject to direct comparison by using the labels and ground truth masks defined by the evaluation methodology proposed for this purpose, by comparing the contours obtained by thresholding the boundaries present between regions after clustering an image with the reference contours present on the dataset. Three out of four methods are based on classical general-purpose clustering methods, namely, K-means, Gaussian Mixture Models and Hierarchical segmentation. Finally, the last method is based on hierarchical watersheds with extended h-minima imposition.

With respect to the image input representation to each of the segmentation methods, it is possible to notice that due the difference between the implementation and purpose of the clustering methods and the morphological segmentation based on watersheds, the input representation may vary, to accomplish this task, the image was evaluated over several colorspaces (RGB, HSV and La*b*), however, to cluster this representation, it is necessary to unroll each pixel coordinate triplet (1) as a column vector, as described on (2), in this case, the spatial coordinates maybe added to enforce spatial segmentation coherence and reduce the number of disjoint regions labeled to the same cluster without being in the same spatial area. Analogically, to process the image by employing the watershed transform, the grayscale gradient was calculated by using the Sobel derivatives, the flooding procedure was done over the image representation on any colorspace, without any spatial information. Finally, the contour detection and thresholding of the segmentation was done using the Canny edge detector.

$$\begin{bmatrix} (R_{1,1}, G_{1,1}, B_{1,1}) & (R_{1,2}, G_{1,2}, B_{1,2}) & \cdots & (R_{1,n}, G_{1,n}, B_{1,n}) \\ (R_{2,1}, G_{2,1}, B_{2,1}) & (R_{2,2}, G_{2,2}, B_{2,2}) & \cdots & (R_{2,n}, G_{2,n}, B_{2,n}) \\ \vdots & \cdots & \ddots & \vdots \\ (R_{m,1}, G_{m,1}, B_{m,1}) & (R_{m,2}, G_{m,2}, B_{m,2}) & \cdots & (R_{m,n}, G_{m,n}, B_{m,n}) \end{bmatrix} \tag{1}$$

$$\left[\begin{array}{cc|c|c} R_{1,1} & R_{1,2} & & R_{m,n} \\ G_{1,1} & G_{1,2} & & G_{m,n} \\ B_{1,1} & B_{1,2} & \cdots & B_{m,n} \\ 1 & 1 & & m \\ 1 & 2 & & n \end{array}\right] \tag{2}$$

The implementation of the watershed procedure is based on the OpenCV[1] Python bindings, with the regional minima extraction routines adapted from the Gala library[2], the clustering methods are based on the Sckit-learn[3] implementations. Also, the centroid and means initialization for K-Means and GMM was based on the KMC[2][4] algorithm [1], all the routines were implemented on Cython, to improve execution times and the efficiency of the evaluation procedure.

[1]http://opencv.org
[2]https://github.com/janelia-flyem/gala
[3]http://scikit-learn.org/stable/
[4]https://github.com/obachem/kmc2

About the models

K-Means

In the present experiment, K-Means was employed to segment an image by conforming clusters from pixel information, the clusters means were calculated by using the ℓ_2 distance metric (Euclidean), which gives rise to spherical and convex color clusters, which can be spatially sparse or localized, depending on the inclusion of spatial information. To segment an input image, for each color triplet, the label of the closest mean is assigned, finally, the segmented image is reshaped into the image original dimensions.

Gaussian Mixture Model (GMM)

The GMM procedure pretends to explain the input image[5] as the set of points spanned by a set of K gaussian distributions. To fit the set of distributions, it is essential to find the maximum likelihood parameters of each gaussian, namely, its mean μ_k and its covariance matrix Σ_k. To maximize those parameters, the Expectation Maximization (EM) algorithm is employed, this routine employs the Mahalanobis distance to model the data, which implies that the distribution may present skewness and doesn't have to be spheric or symmetrical around each axis. The maximization of the a priori probability of the model given the data allows to assign a membership probability to each of the K distributions to each value. In practice, the label of the distribution which have the greater probability is chosen.

Hierarchical Clustering

The hierarchical clustering allows to define segmentation regions by merging smaller pixel regions, which in turn are formed by joining other smaller pixel regions, until each pixel of the image corresponds to its own region. To accomplish this process of merging, it is necessary to define a similarity measure between groups, if the similarity is minimal between clusters, then both are merged and a new cluster is formed. Due to its hierarchical character, this model allows to group several segmentations, which in turn allows to evaluate and pick the best level. However, computing the similarity between clusters is a expensive operation which requires significant amounts of memory, this means that this procedure can't scale if the input dimension increases. During the present experiment, the Ward linkage was used to relate the clusters, this measure expresses that not just only the centroids of both clusters must be near to each other, but each value present in both sets must be near, this measure allows to join spatially close values if the spatial coordinates are included.

[5] It is assumed that its values are normally distributed

To obtain K segmentation regions, the merging algorithm must be repeated until only K groups are left. Due to the excessive memory consumption, each image had to be down scaled by a factor of 0.5, however, this implied a memory consumption of 12Gb, less than the 60Gb required to evaluate the original image.

Watershed Transform

The watershed transform is based on the concept of watershed lines, these lines correspond to the local basin in which all drops of water are gathered when they fall down from a peak. In this context, the peaks of the images correspond to contours, and the space between borders may correspond to a river basin, the idea of the algorithm is to flood the image from the river basins (Regional minima), up to the peaks, at the end of the procedure, only the peaks remain, which implies that a boundary segmentation of the image is achieved. However, due to the existance of multiple regional minima on the gradient, this method is prone to produce an oversegmentation of the image, to overcome this limitation, it is possible to define a set of precomputed markers and impose them over the image, by executing this procedure, the flooding is not done over the set of markers, reducing the oversegmentation on the image. Due to the inexistance of precomputed markers over each image, the markers are selected as the regional minima of height h (h-minima), the value of h is increased, starting on 0, until the total number of regions is equal to K.

3. Results

To evaluate unsupervised methods applied to image segmentation, it was necessary to formulate a common evaluation ground, in which every method can be compared directly with other methods, to address this issue, a subset of eight images were sampled from the Berkeley BSDS Segmentation database, each image have four levels of segmentations, ranked according to the number of regions labeled, also each level contains its contour map (y). The evaluation procedure consists on taking the original image and for each of segmentation levels (K regions), apply any colorspace, include or not include spatial information and then applying a segmentation method to obtain a segmentation that presents K regions, then it is threshold and binarized to obtain a contour map ($\hat{y}$), which is compared to the ground truth employing the mean absolute difference (3)

$$Acc = 1 - \frac{||y - \hat{y}||}{||y + \hat{y}||} \tag{3}$$

This evaluation method presents several shortcomings related to the comparison estimate

However, the color information alone is not enough to get a reasonable segmentation image because this method tends to group distant regions that present a similar color value on any space, this result may be relevant for other applications such as image compression but in the context of image segmentation this may degrade the performance of the method, to account for this limitations, it is possible to include the spatial coordinates to each input vector, however, this causes the appeareance of spherical localized regions, which implies that the final result of the segmentation is close to a Voronoi pattern.

4. Conclusions

References

[1] O. Bachem, M. Lucic, S. H. Hassani, and A. Krause. Approximate k-means++ in sublinear time. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, AAAI'16, pages 1459–1467. AAAI Press, 2016.

[2] S. Beucher. *Watershed, Hierarchical Segmentation and Waterfall Algorithm*, pages 69–76. Springer Netherlands, Dordrecht, 1994.

[3] D. Comaniciu and P. Meer. Mean shift: a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619, May 2002.

[4] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, Aug 2000.