Region Merging for Image Segmentation Based on Unimodality Tests

Theofilos Chamalis, Aristidis Likas
Department of Computer Science and Engineering
University of Ioannina
Ioannina, Greece
e-mail: {thchama, arly}@cs.uoi.gr

***Abstract*—A novel approach is proposed for image segmentation based on region merging. We define a new criterion to decide on whether to merge two regions that does not require the specification of user defined thresholds. The method begins with an image oversegmentation (based on SLIC superpixels) into small homogeneous regions. At each step regions are iteratively merged to form larger regions based on the result of a merge test that measures unimodality as an indication of visual content homogeneity. More specifically, the merge test employs the dip-dist criterion that decides on the unimodality of a set of data objects through the application of the Hardigans' dip-test for unimodality. We propose a fast version of the dip-dist criterion, where the two feature centroids of the regions to be merged are used to decide on the unimodality of the region that results from merging. To demonstrate the performance of the method, we provide segmentation results for several images from well-known image datasets using CIELAB color as the feature vector of each pixel.**

Keywords-computer vision; image segmentation; region merging; dip test for unimodality

I. INTRODUCTION

Image segmentation is the partitioning of an image into disjoint regions called segments that are homogeneous with respect to some visual characteristic. Segmentation is a problem of a particular importance in computer vision and image analysis with numerous applications in robotics, human-computer interfaces, surveillance etc. Due to its importance and wide spectrum of applications, image segmentation has been widely studied for a long time and numerous techniques have been developed [1], [2]. All those categories of methods have advantages and drawbacks and their effectiveness largely depends on the particular characteristics of the image to be segmented.

In this work we present a general unsupervised image segmentation approach that is based on *region merging* starting from an initial large set of small visually homogeneous image segments that are called *superpixels* [3], [4]. However, methods that generate superpixels result in image oversegmentation, thus, there is a need to subsequently merge superpixels with similar visual content. To do this a similarity (or distance) measure between image regions must be defined (eg. based on average color values) and also a threshold-based criterion is needed to decide whether two image regions should be merged or not [5]-[7]. The appropriate specification of user defined thresholds (or other parameters such as the number of segments) largely affects the segmentation result, thus constituting a major weakness for image segmentation techniques.

In this paper, we present an image segmentation approach based on region (superpixel) merging. The merging criterion is based on the idea of measuring the unimodality of a set of data objects as an indication of the content homogeneity in this data set. This idea has been implemented in the dip-dist criterion [8] that decides on the content homogeneity of a set of data objects by applying the dip-test for unimodality [9] on the rows of the distance matrix containing the pairwise distances among the data objects. In our image segmentation case, the data objects are the feature vectors of the pixels, and we propose a modification of the dip-dist criterion that can be applied when merging two unimodal sets of objects in order to decide whether the resulting larger set will be unimodal or not. This modification relies on the centroids of the two sets of objects and requires only two applications of the dip-test for unimodality. This results in significant speedup with respect to the original dip-dist criterion. This speedup is crucial in the case of image segmentation problems where the number of data objects (pixel feature vectors) is large.

In Section II we describe the proposed region merging criterion (called merge test) that is based on a statistical test (dip-test) for unimodality. In Section III we describe the details of the proposed image segmentation technique. In Section IV we provide visual examples from the application of the method while in Section V we provide conclusions and suggestions for further study.

II. THE REGION MERGING CRITERION

We assume that each image pixel is represented by a vector of low level features (e.g. color, texture, PCA features etc). The proposed method relies on a merge test that takes as input a pair of homogeneous image segments (regions) and decides whether the new region that results from merging is homogeneous or not.

To decide on the content homogeneity of the new region we rely on the dip-dist criterion [8] proposed for evaluating the cluster structure of a set of data objects. This criterion is based on Hartigans' dip-test for unimodality [9].

Given a set of real numbers F, the dip-test computes the dip value of F (dip(F)) which is the departure from unimodality of the empirical distribution of F. In other words,

978-1-5090-6086-3/17/$31.00 ©2017 IEEE

the dip value provides a quantitative indication of whether the histogram of a set of real numbers is unimodal (has one peak) or not. It is reasonable to consider that if a set of real numbers has been generated from a unimodal distribution (e.g. a Gaussian) then this set is homogeneous with respect to content, otherwise it is content inhomogeneous. The dip test returns not only the *dip value*, but also a *p-value* indicating the statistical significance of the computed dip value. A p-value equal to zero is a strong indication of multimodality, while a p-value equal to one is a strong indication of unimodality. In this work, we consider that the test decides unimodality if the p-value is greater than zero. For example, the histograms in the second row of Fig. 1 are clearly multimodal, their dip values are 0.054 and 0.043, while their p-values are zero. On the contrary, the histograms in the second row of Fig. 2 are clearly unimodal, their dip values are much smaller (0.012 and 0.014), while their p-values are close to one (0.991 and 0.927 respectively).

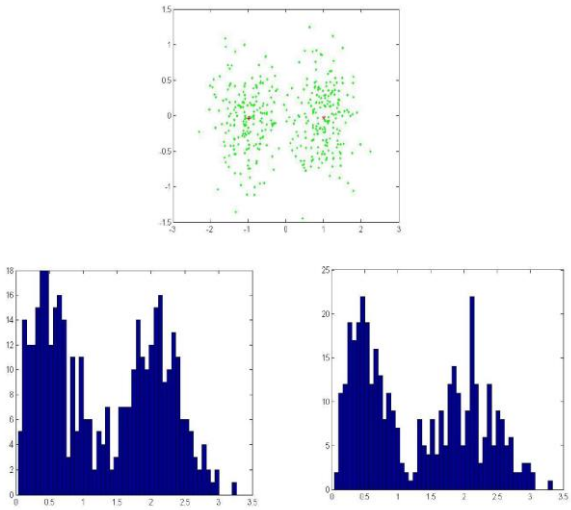

Figure 1. Top: A 2-d dataset consisting of two well-separated groups (green dots). Red crosses correspond to group centroids Bottom: The histograms of the distances in sets D_1 and D_2 respectively. Note that histograms are clearly bimodal.

To decide on the content homogeneity of a group of data objects (pixels), the dip-dist criterion relies on the notion of a viewer which is a data object that decides on the unimodality of the group by considering the set of its pairwise distances from to all other data objects. Then, the density of this set of distances is tested for unimodality using dip-test and the viewer decides either unimodality (unimodal viewer) or multimodality (multimodal viewer). In the original dip-dist criterion all data objects of the group are considered as viewers. If the percentage of viewers suggesting multimodality exceeds a given threshold, then the set of objects is characterized as multimodal, otherwise it is considered unimodal.

In the merge test criterion proposed in this work only two viewers are considered, which are the feature centroids (i.e. average feature values) of the two pixel groups that are tested for merging. More specifically, let R1 and R2 the groups of pixels that are tested for merging and R the group of pixels resulting from the union of R1 and R2. We first compute the corresponding centroids c1 and c2 in the feature space of R1 and R2. Then we form the set D1 containing the distances in feature space between c1 and every pixel in R and the set D2 containing the distances in feature space between c2 and every pixel in R. Next the sets D1 and D2 are checked for unimodality using the dip-test. If both sets are found unimodal, then the region R is considered unimodal and the merge test of R1 and R2 is considered successful. If at least one of the sets D1 or D2 is found multimodal then region R is considered multimodal and the merge test of R1 and R2 fails.

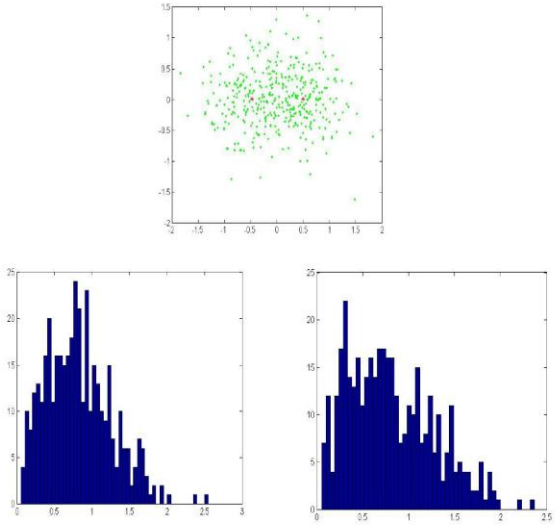

Figure 2. Top: A 2-d dataset consisting of two overlapping groups (green dots). Red crosses correspond to group centroids. Bottom: The histograms of the distances in sets D_1 and D_2 respectively. Note that histograms are clearly unimodal.

In Fig. 1 and Fig. 2 we provide an illustration of the merge test in the case of two dimensional datasets each one consisting of a pair groups (top row of each figure). The centroid of each group is also presented. In the bottom row of each figure, the histograms of the Euclidean distances of each centroid to the whole dataset is given. For the dataset of Fig. 1 the two groups are sufficiently apart, thus they should not be merged, since their union will not result in a compact group. In this case, it can be clearly observed that the two histograms are multimodal. This is also verified by the dip-test providing p-values equal to zero, thus the merge test for this pair of groups fails. For the dataset of Fig. 2 the two groups are sufficiently close, thus they could be merged, since their union results in a compact group. In this case, it can be clearly observed that the two histograms are unimodal. This is also verified by the dip-test providing p-values near 1 (0.991 and 0.927), thus the merge test for this pair of groups is successful.

III. Image Segmentation Based on Region Merging

The proposed image segmentation methodology starts with an oversegmentation of the image into a large number of small homogeneous segments, usually called superpixels.

This can be done using various methods (k-means, normalized cuts etc) [3], [4]. Let R1,…, RM the initial set of image superpixels (e.g. M=150).

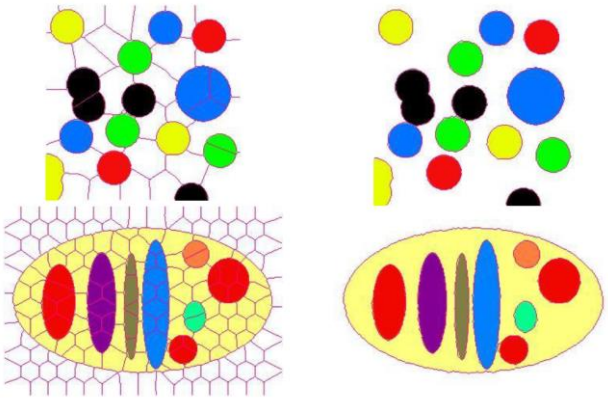

Figure 3. Segmentation results on artificial images. Left column: initial segmentation. Right columns: final segmentation.

Next we proceed in an agglomerative way. At each iteration:

The merge test is applied for all pairs of current regions and the unimodal pairs are first determined, ie. those whose union would result in larger homogeneous regions.

From the unimodal pairs, we determine the pair of regions whose region centroids c1 and c2 in feature space are the closest. These two regions are actually merged and replaced by the new larger region in the current solution.

The above two-step procedure is applied until we reach a solution where all pairs of regions are found to be multimodal according to the merge test. This means that all formed regions are different with respect to visual content, thus their merge would result in visually inhomogeneous regions.

Note that, once the initial oversegmentation is given, the method does not require any parameter to be specified by the user. This is in contrast to most region merging techniques that make use of similarity thresholds to decide whether the merge test is successful or not [5]-[7].

IV. Experiments

The initialization stage of our method requires computing an oversegmentation of the image in a large number of superpixels, which are small, compact and visually homogeneous regions. Based on the comparative results in [10], SLIC superpixels (Simple Linear Iterative Clustering) seem to provide a fast and efficient choice, thus we have used this method in our experiments. SLIC works in the CIELAB color space and performs k-means clustering of the image pixels taking into account the spatial position of the pixels. The approximate number of superpixels is given as input to SLIC (50 in our case), along with parameters related to i) the weighting factor between colour and spatial differences (2 in our case), ii) the radius of the smallest allowed superpixel area (2 in our case).

In the current implementation of our method we consider the CIELAB color representation of an image, thus the feature vector of a pixel is its (L,a,b) vector. The centroid feature vector of a region contains the average (L,a,b) values of its pixels, while the distance between the centroid feature vector and a pixel feature vector is their squared Euclidean distance.

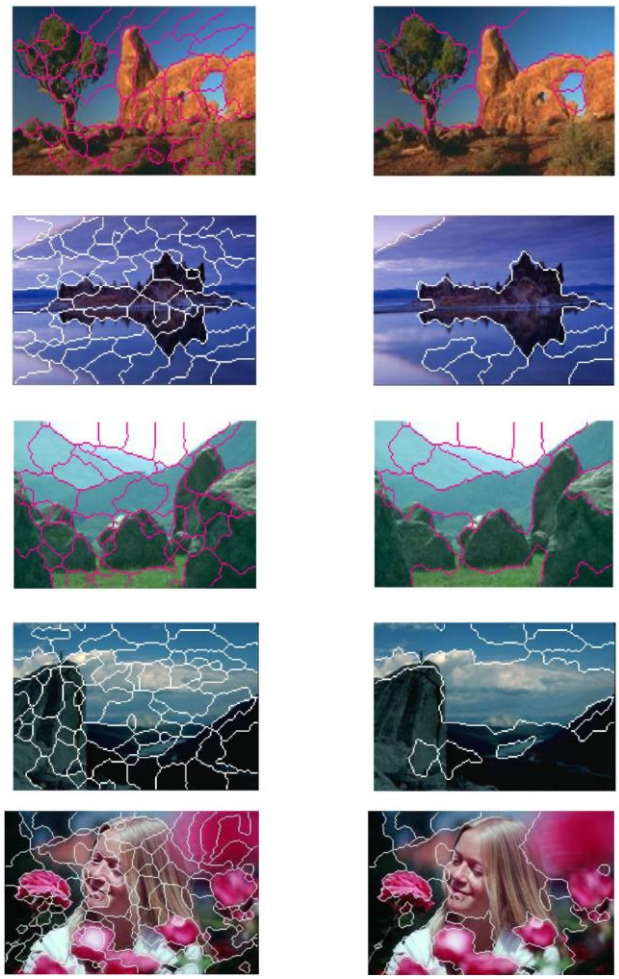

Figure 4. Segmentation results on images from Berkeley image dataset. Left column: initial segmentation. Right column: final segmentation.

At first, in order to test the ability of the method to automatically determine the correct number segments we created artificial images containing distinct objects of various colors. For those images the ground truth segmentation is obvious and unambiguous and the method correctly discovered the true segmentations. Two characteristic examples are presented in Fig. 3. Next we tested the proposed method using images from two well-known image datasets, namely the Berkeley image database [11] and the MSRA image database [12]. Some segmentation examples are shown in Fig. 4 and Fig. 5 respectively.

In all examples the SLIC algorithm for superpixel generation has been run with the same parameter values. As it can be observed, the method is quite successful in merging superpixels of similar visual context despite the fact that i) only color is used as a feature, ii) the method does not require any parameter to be specified (except for the initial

SLIC parameters). It must be noted that the method in its current form proceeds through merging superpixels thus, it cannot recover from a bad initial segmentation, where there exist single superpixels crossing the true segment boundaries. For this reason, it is important that the initial segmentation respects the ground truth region boundaries.

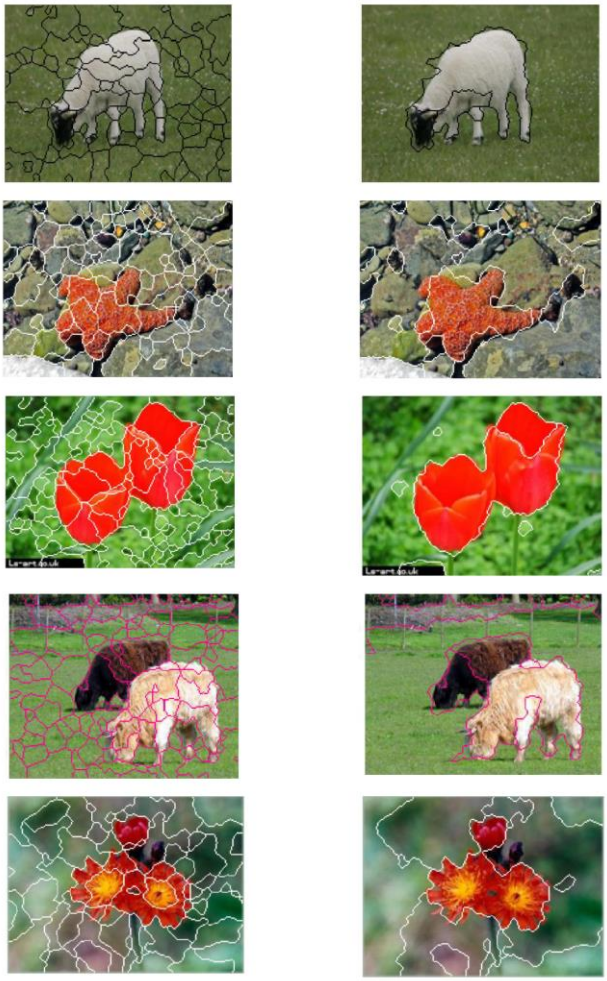

Figure 5. Segmentation results on images from the MRSA dataset. Left column: initial segmentation. Right column: final segmentation.

V. CONCLUSIONS

In this paper we have proposed a novel approach for image segmentation based on superpixel merging that automatically estimates the final number of segments without requiring the specification of threshold value on the similarity between superpixels.

The method follows the idea of agglomerative clustering, and at each step, starting from an initial segmentation based on SLIC superpixels, regions are iteratively merged to form larger unimodal regions based on the result of a merge test that employs Hardigans' dip-test for unimodality. Segmentation results on several images using color as feature indicate that our method provides segmentations of acceptable quality without requiring the specification of similarity thresholds.

There is ample room for further investigation on the proposed approach. Although color has been used in this study, other types of features could have been examined, i.e. texture features, PCA features, edge maps etc. Moreover, a quantitative performance evaluation of the method is planned by comparing the obtained segmentations to the human segmentations available with several image datasets.

REFERENCES

[1] C. Russ, The Image Processing Handbook, 6th edition, CRC Press, 2016, ch. 7, pp. 395-443.

[2] A. Mitiche and I. Ben Ayed, Variational and Level Set Methods in Image Segmentation, Springer, 2010.

[3] A. Moore, S. Prince, J. Warrell, U. Mohammed, and G. Jones, "Superpixel lattices," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (ICCV 08), 2008, pp. 1-8.

[4] B. Fulkerson, A. Vedaldi and S. Soatto, "Class segmentation and object localization with superpixel neighborhoods," in Proc. 12th IEEE Int. Conf. Computer Vision (ICCV 09), 2009, pp. 670-677.

[5] C.-Y. Hsu and J.-J. Ding, "Efficient Image Segmentation Algorithm Using SLIC Superpixels and Boundary-focused Region Merging," Proc. 9th Int. Conf. on Inform., Comm. & Signal Proc., 2013.

[6] F. Nielsen and R. Nock, "Consensus Region Merging for Image Segmentation," Proc. 2nd IAPR Asian Conference on Pattern Recognition (ACPR 13), 2013.

[7] L. Li, J. Yao, J. Tu, X. Lu, K. Li and Y. Liu, "Edge-Based Split-and-Merge Superpixel Segmentation," Proc. IEEE Int. Conf. on Information and Automation, 2015.

[8] A. Kalogeratos and A. Likas, "Dip-means: an incremental clustering method for estimating the number of clusters," Proc. Neural Information Processing Systems (NIPS 12), 2012.

[9] J. A. Hartigan and P. M. Hartigan, "The dip test of unimodality," Annals of Statistics, vol. 13, no. 1, 1985, pp. 70-84.

[10] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua and S. Susstrunk, "SLIC Superpixels Compared to State-of-the-art Superpixel Methods," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, no. 11, pp. 2274-2282, May 2012.

[11] 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," Proc. 8th Int. Conf. Comp. Vision (ICCV 01), 2001, vol. 2, pp. 416—423.

[12] T. Liu, J. Sun, N.-N. Zheng, X. Tang and H.-Y. Shum, "Learning to Detect A Salient Object," Proc. IEEE Conf. on Comp. Vision and Patt. Rec. (CVPR 07), 2007.